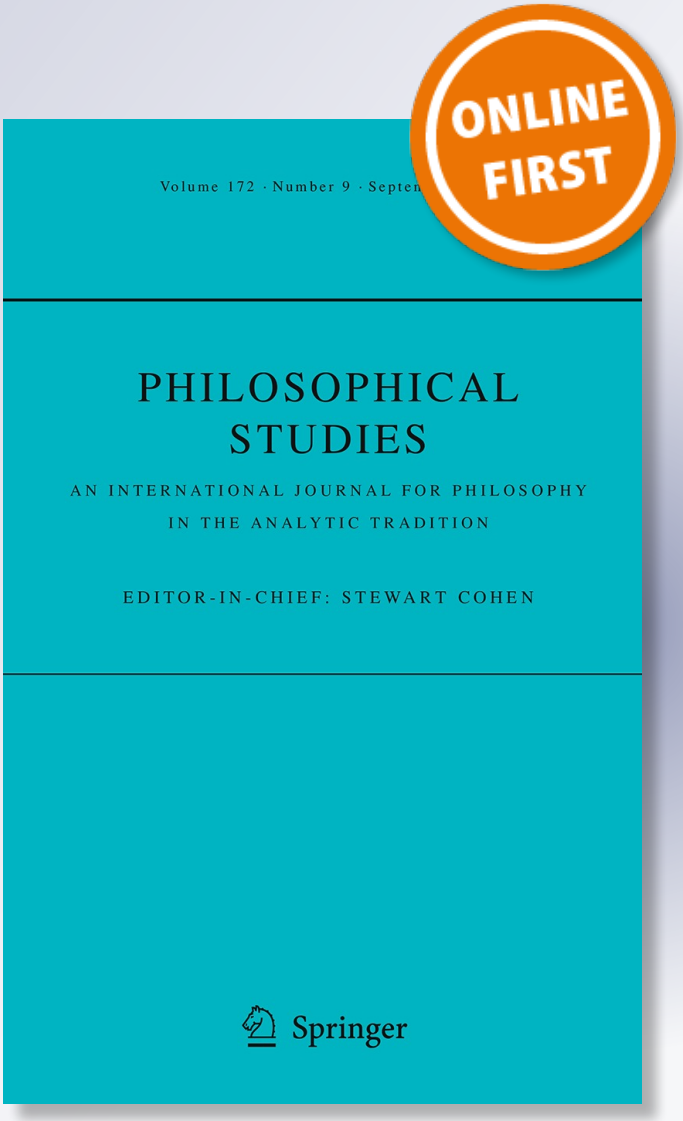
Serious theories and skeptical theories: Why you are probably not a brain in a vat

Michael Huemer

Philosophical Studies
An International Journal for Philosophy in the Analytic Tradition

ISSN 0031-8116

Philos Stud
DOI 10.1007/s11098-015-0539-5



Springer

Philos Stud
DOI 10.1007/s11098-015-0539-5

Serious theories and skeptical theories: Why you are probably not a brain in a vat

Michael Huemer[1]

Abstract Skeptical hypotheses such as the brain-in-a-vat hypothesis provide extremely poor explanations for our sensory experiences. Because these scenarios accommodate virtually any possible set of evidence, the probability of any given set of evidence on the skeptical scenario is near zero; hence, on Bayesian grounds, the scenario is not well supported by the evidence. By contrast, serious theories make reasonably specific predictions about the evidence and are then well supported when these predictions are satisfied.

Keywords Skepticism · Cartesian skepticism · Brain in a vat · Bayesianism

1 The argument from skeptical scenarios

1.1 The irrefutability of skeptical scenarios

A skeptical scenario is a possible situation in which everything might appear to us the way it does now, but most of our beliefs in some broad area would be false. For instance, I could be a brain in a vat, and the scientists stimulating my brain might be giving me experiences just like those of a normal person moving about the world, when in reality none of what I seem to perceive is real. Or God could be planting false images of physical objects in my mind, when in reality there is no physical world.[1]

[1] For a brain-in-a-vat scenario, see Pollock and Cruz (1999, pp. 2–5). For the deceiving God scenario, see Descartes (1984, p. 14).

✉ Michael Huemer
bmsjrqcrna@snkmail.com

[1] Philosophy Department, CB 232, University of Colorado, Boulder, CO 80309-0232, USA

Published online: 22 August 2015

Skeptical scenarios are notoriously difficult to refute. This is because the scenarios, such as the brain in the vat, the deceiving God, or the dream hypothesis, are generally capable of accommodating any experience. I cannot conduct an empirical test of whether I am a brain in a vat, because any apparent results of any empirical test could be explained by the clever scientists with their brain-manipulating supercomputer.

Or so the skeptic would claim. What I think is true is that there is no possible series of sensory experiences that a skeptic would accept as sufficient grounds for rejecting the BIV scenario. To the extent that one thinks the BIV scenario explains our actual experiences, it could explain any other experiences about equally well or better. My contention in what follows, however, will be that the BIV scenario provides an extremely poor explanation of sensory experience.

1.2 The target skeptic

I am not here interested in a skeptic who says that merely because there is a nonzero (perhaps incredibly tiny) probability that one is a brain in a vat, one does not "know" facts about the external world. Nor am I concerned with the a priori skeptic, who maintains that there is no a priori justification, or the global skeptic, who maintains that there is no justification for any belief.

The skepticism with which I am concerned is external-world, justification skepticism. My skeptic holds that because we have no reason to reject the brain-in-a-vat scenario, or at least no strong reason to reject it, our beliefs about the external world are not even *justified*. This is an interesting conclusion: it is very surprising, and yet there is a prima facie plausible argument for it.

1.3 The skeptic's argument

Let "BIVH" stand for the hypothesis that I am a brain in a vat whose experiences are caused by scientists with a supercomputer in the manner usually envisioned. Let "RWH" stand for the contrary hypothesis that I am a normal person perceiving the real world. The skeptic would have me conclude that, since I can never refute the BIV scenario, I can never justifiably believe RWH. Here is my best reading of the argument:

1. I am justified in believing RWH only if I am justified in rejecting BIVH (premise).
2. I am justified in rejecting BIVH only if (a) I have a priori justification for rejecting BIVH, or (b) my sensory experiences provide evidence against BIVH (premise).
3. I have no a priori justification for rejecting BIVH (premise).
4. BIVH predicts that my sensory experiences would be just as they are (premise).[2]

[2] How should the skeptic understand the notion of "prediction" here? One interpretation is entailment: the BIV scenario could be described such that it just entails that one has the sort of experiences one in fact

5. For any evidence E and hypothesis H, if H predicts E, then E does not provide evidence against H (premise).
6. My sensory experiences do not provide evidence against BIVH (from 4, 5).
7. I am not justified in rejecting BIVH (from 2, 3, 6).
8. I am not justified in believing RWH (from 1, 7).

There are many points where this argument could be challenged and has been challenged.[3] But I'm going to ignore most of them. There is just one response that I want to discuss here. This response claims that the BIV scenario, while *consistent* with our actual sensory experiences, is *highly improbable* on those experiences. Exactly which step of the above argument is wrong depends upon how we construe BIVH. On one construal, it will turn out that premise (4) is false because the probability of our having the sort of experiences we do *given* BIVH is very low. On another construal, premise (3) is false because BIVH has a very low a priori probability, and this is a reason for rejecting it.[4]

2 A conception of probability

2.1 Philosophical assumptions

My antiskeptical argument employs a certain conception of probability. I can't present a full defense of this conception here since there are very large and complex issues in this area, although I shall touch on some common objections in Sect. 5. For now, I will simply lay down the view, as follows.

There is a kind of probability, "logical probability," which satisfies the standard axioms of probability theory and which directly bears on epistemological questions. Of particular interest here, both RWH and BIVH have a logical probability conditional on our sensory experiences. If the probability of RWH given our sensory experiences is greater than that of BIVH given our sensory experiences, then we have some epistemic reason to prefer RWH over BIVH (and also some reason to reject BIVH). We have some intuitive insight into facts about probability, including not only the standard Kolmogorov axioms but also certain assessments of the probabilities of particular contingent propositions. This includes our intuitions about *at least some* applications of the Principle of Indifference. That is to say, in some cases, where we intuitively see that a particular partition of the possibilities is

Footnote 2 continued
has. A more flexible interpretation (which leaves both premises 4 and 5 plausible) would be probabilistic: H predicts E provided that $P(E|H) > P(E)$.

[3] For a sample of approaches, see Putnam (1981), Vogel (1990), Dretske (1981), Klein (1995). In Huemer (2001a, pp. 388–389), I provide grounds for rejecting premise 5.

[4] For a brief, early defense of this approach, see BonJour (1985, pp. 183–185). Note that my present defense of the approach does not indicate an abandonment of the direct realist response to skepticism presented in my earlier work (Huemer 2000, 2001b, pp. 181–191); rather, I believe the unsoundness of the skeptical argument is overdetermined.

especially natural, and we lack reasons for favoring any possibility over any other, we can also see that each possibility has an equal initial probability.

The rest of this paper can be read as defending the conditional claim that if one accepts this view of probability, then one has available a strong rebuttal to Cartesian skepticism. This conditional is interesting in part because a number of interesting thinkers have taken conceptions of probability in this neighborhood very seriously.[5]

2.2 Bayes' theorem

Bayes' Theorem tells us, for any hypothesis H and evidence E:

$$P(H|E) = \frac{P(H) \times P(E|H)}{P(E)}.$$

According to the logical interpretation, the posterior probability of H, P(H|E), represents the degree of support H has, when E is one's total relevant evidence. Three factors affect this: the prior (a priori) probability of H, P(H); the prior probability of E, P(E); and the probability of E given H, P(E|H). P(E|H) is also sometimes confusingly called the "likelihood of H" (confusingly because it is a probability *of* E, not of H).

2.3 Assessing likelihoods

For the moment, leave aside the priors and focus on the likelihood factor. The higher the likelihood P(E|H) is, the higher will be the posterior probability of H. Intuitively, we could describe P(E|H) as the degree to which H *predicts* E. If H *entails* E, then P(E|H) will take on the maximum value, 1. If H *precludes* E, then P(E|H) will take on the minimum value, 0.

Between those two extremes, H may generate stronger or weaker expectations as to whether E would be the case. How can we assess these expectations in the intermediate cases? One possibility is that we simply make an intuitive guess—imagining H as being the case, we find that it seems more or less plausible that E would hold. But in some cases, we lack any intuitive sense of how likely E would be, and in any case, these intuitive guesses are likely to suffer from low reliability.

A second approach is to apply the Principle of Indifference to the possible ways the evidence could be. If H is compatible with a number of different possible sets of evidence, say $\{E_1, E_2, E_3, \ldots, E_n\}$, and if we have no reason to think that H would favor any one of these alternatives over any of the others, we could assign probability $1/n$ to each E_i. This approach requires that there be a particular way of dividing up the alternatives that is especially natural. Often, there is such a way; if not, one must take another approach.

A third approach is to imagine further specifications of the hypothesis—that is, ways of filling in details that are not specified by H, or specifying the values of certain parameters in H—where these added specifications would enable us to have

[5] Keynes (1921), Carnap (1962), Jaynes (2003), Fumerton (2004), Huemer (2009).

an intuitive sense of whether E would hold. Thus, suppose that H is compatible with any of H_1, H_2,…, H_n, where this represents a natural partition of the possibilities. Suppose we firmly judge that if any of H_1,…, H_m held, then E would (certainly or almost certainly) hold, and if any of H_{m+1},…, H_n held, then E would (certainly or almost certainly) not hold. Suppose finally that we have no reason for favoring any of the H_i over any of the others. Then, with the help of the Principle of Indifference, we can approximate the value of P(E|H) as m/n.

What if none of these approaches work? We have no intuitions about the likelihood, there is no especially natural way of dividing up the alternative evidences, and there is no especially natural way of dividing up the alternative ways of filling in the hypothesis so as to generate clear predictions regarding E. In that case, we may be unable to figure out P(E|H). Let's hope that doesn't happen too often.

In Sect. 5 below, I will address skeptics who doubt that there are any a priori probabilities, or that we can ever reliably assess them. But first I want to explain the basic argument for the improbability of the brain-in-a-vat scenario.

3 The broad skeptical scenario is disconfirmed by experience

3.1 The broad BIV hypothesis

There are at least two ways of interpreting the BIV hypothesis, a broad interpretation and a narrow interpretation. On the *broad* interpretation, the hypothesis posits a certain mechanism by which my experiences are generated (a scientist with a computer stimulating a brain in a vat) but does not specify anything about what sort of experiences that mechanism generates. Thus, the *Broad BIVH* (as I shall call it) is compatible with my having any sequence of experiences that a brain could be induced to have. By contrast, on the *narrow* interpretation, the BIV hypothesis specifies not only the source of my experiences but also something about their content or about the intentions of the scientists. In other words, the Broad BIVH says, roughly, "I am a brain in a vat who is being stimulated so as to have some experiences," whereas the Narrow BIVH says something like, "I am a brain in a vat who is being stimulated so as to have *experiences like those of a normal human being*."

In this section, I focus solely on the Broad BIVH. I contend that the Broad BIVH is a bad theory because it has a very low likelihood; relatedly, premise 4 in the skeptic's argument—

4. BIVH predicts that my sensory experiences would be just as they are.

—is false on the broad interpretation.

What does the Broad BIVH predict about our experiences? Almost nothing. Given only the information that a brain in a vat is hooked up to a computer and is being stimulated so as to have experiences, any sequence of experiences is possible. Perhaps the brain would be having experiences like those of a normal human being. Then again, perhaps the brain would be having a completely random and incoherent series of experiences. Either would be equally within the powers of the computer

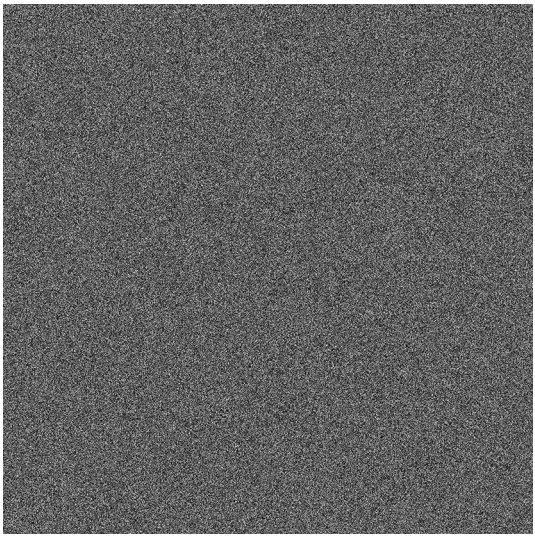


Fig. 1 Image generated by placement of randomly colored pixels

stimulating the brain. How should we assess the probability that the brain's experiences would be like ours?

3.2 Applying the principle of indifference to the possible evidence

Suppose we applied the Principle of Indifference, assigning an equal probability to every possible sequence of experiences that the computer could stimulate the brain to have. Then, with near 100 % certainty, the brain would have utterly disordered, uninterpretable experiences. To see what I mean by this, start with a toy example. Imagine a video display 1000 pixels wide and 1000 pixels tall. Now imagine a program that produces an image by assigning a random color to every pixel. What would the image look like?

Next, imagine that the program uses the same process to generate an image fifty times every second. Suppose you viewed these images for a period of 1 minute as they played across the screen. What would this movie look like?

The answer to the first question is provided by Fig. 1, which is a black-and-white version of an image generated by the process described above; that is, a computer was programmed to place a random color at each position in a 1000 × 1000 grid. I used the same program to generate 200 such images, but it is only necessary to reproduce one of them here, since all of them look alike. Although each image is a distinct arrangement of colored pixels, completely independent of every other image, every one of them is essentially similar to Fig. 1, as far as any normal human observer is concerned. If one views these images in rapid succession, the result is equally meaningless.[6]

[6] The program assigned a random HSV combination to each pixel, from among 16.7 million possibilities. My computer is unable to display 50 of these per second, but a slide show playing ten images per second

Object lesson: almost all possible distributions of colors fail to generate anything recognizable as a depiction of objects, and almost all possible sequences of images fail to generate anything recognizable as a depiction of events. Virtually all sequences of images just look like static with no discernible patterns.

The human perceptual field is not a computer screen. Nevertheless, essentially the same point applies. Of all the possible sequences of qualia that a mind could experience, *almost all* would be uninterpretable, in the way that Fig. 1 is uninterpretable. Thus, if we start from just the idea of a mechanism capable of producing any experiences, and we apply the Principle of Indifference, the probability of getting a series of experiences exhibiting coherent patterns is essentially nil. Since our experience in fact exhibits coherent patterns, the Broad BIVH provides an extremely poor explanation of our experiences.

What is meant here by a "coherent pattern"? We needn't take any very strong interpretation of coherence. When it comes to visible images, we can distinguish intuitively between images that *look like something* (that is, images that seem to be images of some thing or things other than themselves) and images that look like random noise, like Fig. 1. The computer program described above generates images that look like random noise. Though it is of course theoretically possible that it should, by chance, generate an image that looks like a photograph of the Eiffel Tower, if you were actually to sit in front of the computer watching the images it generates, I have no doubt that you could watch for the rest of your life without ever seeing an image that looked like anything other than noise.

Of course, our experience of the world on the whole exhibits stronger and more interesting forms of coherence than the minimal kind of coherence possessed by any image that merely "looks like something." But there is no need to attempt to analyze any strong types of coherence, nor is there, indeed, any need to further analyze the weak form of coherence that I have described, because images like Fig. 1 *fail utterly* to exhibit *any* sort of coherence that any epistemologist would be interested in. Perhaps if Fig. 1 appeared to exhibit some kinds of coherence but not others, or if it were a borderline case of some interesting sort of coherence, then we would wish to further analyze candidate notions of coherence. But in fact, Fig. 1 is not anywhere close to exhibiting any interesting form of coherence. Similarly, neither are the overwhelming majority of the logically possible sequences of sensory experiences that a subject could have.

3.3 The characteristics of the scientists

The reasoning of the preceding subsection applies the Principle of Indifference at the level of *possible sequences of experiences* that could be induced in a brain. The skeptic might argue that this is the wrong way of assessing P(*E*|BIVH). Instead, the skeptic might say, we should consider the possible intentions and capabilities of the scientists, because it is these things that will *explain why* the scientists induce one

Footnote 6 continued
shows no detectable patterns. The viewer can tell that the image is changing, but one cannot recognize any specific image when it is repeated.

series of experiences or another in the brain.[7] If the scientists have the goal of giving the brain a coherent set of experiences, then the brain will have a coherent set of experiences, however small the proportion of possible experiences that are coherent may be.

This suggests that we take the approach of imagining different ways of specifying the details of the brain in a vat world—in particular, the characteristics of the scientists and their brain-manipulation equipment—and try to guess what proportion of these specifications would result in the brain's having experiences like those of a normal person. This is of course very difficult, and I shall not attempt a numerical estimate. However, I will argue that on the vast majority of possible specifications, the brain would *not* have experiences like those we actually have. For the brain to have experiences indistinguishable from those of a normal person, the scientists and their equipment would have to have certain very specific characteristics:

1. *The purposes of the scientists:* It is often said that the BIV might have experiences like ours because the scientists decided to program a realistic simulation of life in a primitive society (one that lacks BIV technology). This is a possible purpose for a group of scientists with a BIV. But there are many other possible purposes. The scientists could instead have the purpose of producing pleasure. Or producing the sensation of purple. Or producing an intellectually interesting simulation. Or an artistic simulation. Or a funny simulation. Or a simulation that helps develop moral virtue. Or a simulation of what life might be like in their (the scientists') future. And so on. Many of these other possible purposes seem to make a fair bit more sense than the purpose of producing a perfectly realistic simulation of a bland, ordinary life in the early twenty-first century. And each of them would generate certain expectations about the BIV's experiences that are violated by our actual experience—for instance, that its experiences should be incredibly pleasurable, incredibly interesting, etc. Only a very narrow range of possible scientist purposes would result in the brain's having experiences just like those of a normal person.
2. *The capabilities of the apparatus:* The BIV would have the same sort of experiences as a normal person only if the computer and the rest of the apparatus for stimulating the brain were incredibly advanced. There are many possible lesser levels of technological sophistication such that the equipment periodically suffers hardware failures, there are noticeable delays created by the computer's need to perform calculations, the brain tries to explore a part of the virtual world that hasn't yet been filled in, and so on. The case where the equipment is, as far as the subject can tell, essentially perfect, is only a very narrow range of the possible ways the equipment could be.
3. *The skills of the scientists:* The scientists would have to be unbelievably skillful computer programmers, such that, in a program vastly more complicated than any other program you've ever imagined (how many billions of lines of code would be required to program a perfectly realistic simulation of life?), no bugs

[7] In Huemer (2009), I argue that when possible, the Principle of Indifference should be applied at a more explanatorily basic level, rather than a more superficial level.

> are ever observed. (Aside: this author can't even write a hundred lines of code without generating bugs.) This level of programming skill includes only a tiny portion of the range of possible levels of skill.

Conclusion: if I were a BIV, my experience would almost certainly *not* be indistinguishable from that of a normal person.

3.4 The predictions of RWH

The preceding conclusion invites the following reply from the skeptic:

> Suppose we consider a "Broad RWH," which specifies that your experiences are reliable perceptions of the external world, but it specifies nothing further about the content of these experiences. The probability of your having precisely the experiences you actually have, given only Broad RWH, is extremely low. After all, consider how many other possible sequences of experiences a being with reliable perception could have—you could have had experiences of living on Mars, or of being an octopus, or of living in the 26th century, etc. Thus, P(*E*|Broad RWH) is extremely low, *just as* P(*E*|Broad BIVH) is extremely low.

Indeed, if E is an exact specification of my sensory experiences, $P(E|H)$ will be extremely low for *any* hypothesis that is framed in reasonably broad terms, i.e., that does not contain very precise specifications of a large number of parameter values. But it is important to notice that in arguing that the Broad BIVH has a low likelihood, I did not appeal to very specific details of my experiences—for instance, that I am presently having an experience as of a table with a yellowish tablecloth. Any broad theory is going to have a low likelihood given specific details of that sort; the Broad BIVH and the Broad RWH are in the same boat there. In arguing for the low likelihood of the Broad BIVH, rather than appealing to such specifics, I appealed to certain broad features of our sensory experience, for instance, the fact that our experience can be easily interpreted as representing lasting physical objects, that we do not experience any things that look like program bugs, that we do not experience delays of the sort that might be due to limitations on computing power. The Broad RWH can explain these broad features, while the Broad BIVH cannot. So the Broad BIVH remains in a much worse position.

Here is another skeptical objection:

> The Broad RWH does not really predict the coherence of our sensory experiences. The Broad RWH says only that one is reliably perceiving the physical world, but the physical world could have been constantly changing in unpredictable ways, with properties randomly distributed through spacetime, such that a completely veridical series of perceptions would seem like meaningless noise.

The skeptic is right about this. To explain the coherence of our experiences, we have to allow RWH to include the postulate that we live in a world of lasting physical

objects that obey consistent laws of nature (which is still a very broad claim). But notice that BIVH, too, embraces this assumption. The BIV world contains a brain, a vat, some scientists, a computer, some equipment for stimulating the brain, and so on. These are all lasting physical objects obeying laws of nature, and BIVH needs these objects in order to explain our sensory experiences. This is not to say that BIVH must make any very specific claims about the character of the laws or the rest of the physical world; BIVH need not assume that the real laws of nature are the same as those in the virtual world, nor that the real laws are deterministic, etc. But nor need the Broad RWH specify these things. If we are only trying to explain very broad features of experience, such as the fact that our experience is coherent, RWH need only say that the world contains lasting physical objects subject to laws of nature. By contrast, even after we posit lasting physical objects subject to laws of nature, BIVH *still* fails to predict a coherent series of experiences, since it holds that our experiences do not correspond to any of those objects.

If we appropriately match up BIVH and RWH, giving them the same evidence to explain (e.g., the general coherence of experience), and letting the two theories make the same sort of broad postulates about the world (lasting physical objects subject to laws), RWH gets a much higher likelihood. That is enough for us to have a strong reason to prefer RWH over BIVH.

4 The narrow skeptical scenario is a priori improbable

The skeptic might seek to augment BIVH's predictive power by including some further stipulations about how the BIV world works. There are three obvious ways of doing this:

(a) We could let BIVH include the stipulation that the brain is stimulated so as to have a coherent series of experiences (without specifying any further details about the experiences). In that case, the hypothesis predicts with probability 1 that the brain would have coherent experiences.
(b) We could let BIVH include a description of the brain's specific experiences. BIVH would specify that I am a brain in a vat who is being stimulated to have exactly this set of experiences, including a visual experience of a table with a yellow tablecloth, an auditory experience of a running river, and so on.
(c) We could let BIVH include a specification of certain intentions and abilities of the scientists. Thus, the Narrow BIVH might state that I am a BIV who is being stimulated by a computer with certain extremely advanced capabilities, programmed by scientists with perfect programming skills who decided to create a perfect simulation of a bland life around the turn of the twenty-first century.

Call these three theories "BIVH-a," "BIVH-b," and "BIVH-c," respectively. Any of these approaches would enable BIVH to predict some of the aspects of our experience that I cited in Sect. 3 as things left unexplained by the Broad BIVH.

Nevertheless, these moves make no genuine progress; they just push the problem around, like the proverbial bump in the rug. Consider BIVH-a. In Sect. 3, I argued

that given the Broad BIVH, it is improbable that I should have coherent experiences. Obviously, the problem for BIVH is not defeated if we simply fold that improbability into the theory, letting our new theory be the *conjunction* of the Broad BIVH with the stipulation that the brain has coherent experiences. All we do is trade in an argument for assigning a low likelihood to the theory, for an equally forceful argument for assigning a low prior: Let "COH" be the proposition that we have coherent experiences. BIVH-a is just (Broad BIVH & COH). Therefore, P(BIVH-a) = P(Broad BIVH) × P(COH|Broad BIVH). And therefore, if we have an argument that P(COH|Broad BIVH) is low, P(BIVH-a) will be low for precisely the same reason. Parallel points apply to BIVH-b and BIVH-c. On this account, then, premise 3 in the skeptic's argument—

3. I have no a priori justification for rejecting BIVH.

—would be false. We have a priori grounds for rejecting Narrow BIVH because Narrow BIVH has a very low a priori probability.

5 Questions and objections

5.1 Is improbability a reason for rejection?

I have just assumed that a proposition's having a very low a priori probability would constitute justification for rejecting it.[8] Lottery cases cast doubt on this assumption. Suppose T is a randomly chosen ticket in a lottery with a large number of tickets, in which you know that exactly one ticket will win. On this evidence, it is highly improbable that T will win. But some intuitively judge that one is not justified in believing that T will lose. This seems to show that the mere fact that a proposition is highly improbable on one's evidence does not suffice for one to be justified in rejecting the proposition. Similarly, the skeptic might claim that the mere fact that Narrow BIVH is highly improbable does not constitute justification for rejecting Narrow BIVH.[9]

There is one version of this objection that I simply want to sidestep as beyond the scope of this paper. Specifically, the skeptic might be an infallibilist about justification, holding that "justified rejection" of BIVH entails absolutely conclusive evidence against BIVH, evidence on which the probability of BIVH would be zero. As suggested in Sect. 1.2 above, this kind of skepticism is not the target of this paper. In fact, I agree that BIVH has a nonzero probability on our evidence. I am not here interested in debating whether that means that we are not "justified" in rejecting BIVH or that we do not "know" that ∼BIVH. Rather, I simply assume that justification does not in general require absolutely conclusive evidence.

[8] Or perhaps the low a priori probability would merely be a symptom of our having justification for rejecting the proposition, with the considerations that explain why the a priori probability is so low constituting the justification. This distinction is immaterial for present purposes, so I shall hereafter elide the distinction.

[9] I thank Daniel Singer for a version of this objection.

So, for the lottery case to pose an interesting challenge (or at least one that I am here interested in addressing), the claim must not be that inconclusive grounds never constitute adequate justification. The claim must be that *in lottery cases*, inconclusive grounds do not suffice for justified belief, even though inconclusive grounds may suffice for justified belief in other cases. For example, intuitively, I am justified in believing that my car is in the parking lot outside the building (where I vividly remember parking it a short time ago), even if I am not justified in believing that ticket T is a loser in the lottery case. This is true despite the fact that the probability of T being a loser might be higher than the probability of my car being where I remember parking it.

What is the difference between lottery cases and other cases in which one has inconclusive grounds for a proposition? Here is one possible account: in the lottery case, one has a certain reason for believing "T is a loser," where one also knows (1) that there is some ticket, T′, which is a winner, and (2) that one has the same grounds for thinking T′ is a loser as one has for thinking T is a loser. More generally, in the lottery case, one has certain grounds for believing that P, where one also knows that these same grounds (or *very* closely parallel grounds) support a false proposition.

If something like that explains what is special about lottery cases, then the BIV case is clearly not analogous. It is not the case that we know, or even have good reason to suspect, that the reason we have for denying Narrow BIVH (or some very closely parallel reason) also supports a false proposition. It is not, for instance, as though we know that we are in some skeptical scenario or other, and we only reject Narrow BIVH because there are so many other skeptical scenarios that might be the correct one. *That* would make our situation genuinely parallel to the lottery case.

A rival account of what is special about lottery cases is that in lottery cases, our evidence for the lottery proposition (e.g., "T is a loser") is "merely probabilistic," as opposed to some other, "qualitative" sort of evidence. Very briefly, I think that it is going to be difficult to distinguish "qualitative" evidence from "merely probabilistic" evidence, and I think this view is going to require us to exclude too many ordinary cases of seemingly justified beliefs from counting as justified. For reasons of space, I will have to leave the discussion there for now.

5.2 What is the prior probability of RWH?

So far, I have argued that Broad BIVH has a much lower likelihood than RWH, and Narrow BIVH has an extremely low prior probability. But these claims alone do not suffice to compare RWH to BIVH; in addition, must we not have some assessment of the prior (a priori) probability of RWH?[10]

I have very little idea what the a priori probability of RWH is. Fortunately, we need not assess that probability per se, since we are only seeking a reason to *prefer* RWH *over* BIVH. We can thus proceed in two stages.

[10] Cf. Vogel's (2010, pp. 416–417) objection to BonJour.

First, we can compare the prior probability of the Broad RWH to the prior probability of the Broad BIVH. Whatever the priors of these theories might be, there seems to be no reason for regarding BIVH as having a vastly higher prior than RWH. RWH is not, for example, vastly more complex than BIVH; it does not contain commitments to much more specific claims than BIVH; and it is not much less intuitive than BIVH. The two theories have about equal degrees of simplicity and generality, and the RWH is, if anything, more intuitive than BIVH. Given that Broad BIVH has a vastly lower likelihood than RWH, this suffices to conclude that RWH has much higher posterior probability on the evidence.

Second, we can compare the prior probability of the Narrow BIVH to the prior probability of a parallel Narrow RWH. Start with BIVH-a. A parallel Narrow RWH, RWH-a, would conjoin the Broad RWH with the stipulation that I have coherent experiences. Our argument for assigning a low prior to BIVH-a depended on the claim that P(COH|Broad BIVH) is low, and that BIVH-a is just the conjunction of Broad BIVH with COH. The Narrow RWH, similarly, is the conjunction of Broad RWH with COH; however P(COH|Broad RWH) is *not* nearly as low as P(COH|Broad BIVH). Therefore, there is no reason for assigning a low prior to RWH-a parallel to the reason for assigning a low prior to BIVH-a. Accordingly, we should judge RWH-a as having a higher prior than BIVH-a. Similar points apply in the case of RWH-b and RWH-c.

5.3 Are there a priori, logical probabilities?

Many philosophers, even those who are not skeptical about the external world and not skeptical about probability per se, are skeptical of the notion of a priori, logical probabilities.[11] These thinkers usually doubt that we have any a priori intuitions about probability, and they hold that knowledge of probabilities is always empirical. For example, the first time one sees a coin, these philosophers would say, one will have no intuitions at all about how likely it is that throwing the coin will result in its landing with heads up. Our actual knowledge that this probability is 50 % is based upon long experience with coins.

To these philosophers, I have three observations to make. First, to hold that there are a priori logical probabilities (logical probabilities to which we have some a priori access), one need not hold that for any arbitrary proposition, upon merely considering that proposition, one will have a relevant intuition about its logical probability. One need only hold that we have *some* relevant sources of a priori justification, from which substantive information about probabilities can sometimes be recovered.

Second, I believe it plausible to hold that there are two such sources of a priori justification. One source consists of intuitions about probability, most of which concern general principles, such as the Kolmogorov axioms and the Principle of Indifference, that constrain our assignment of probabilities. Another source is our capacity for reasoning, including especially non-demonstrative reasoning (by far the most common sort of reasoning in which we engage). I maintain that all reasoning

[11] Ramsey (1931, pp. 160–167) and Gillies (2000, pp. 52–53) reject the notion, while Fumerton (1995, p. 218) and Beebe (2009, pp. 628–632) express doubts.

involves mental representations at least implicitly about logical relations. This is because a person counts as genuinely inferring a conclusion from a premise only if that person *perceives* that premise *as supporting* that conclusion. Because most actual cases of reasoning are non-demonstrative, the support relations that we represent will usually be probabilistic. And these representations must be justification-conferring, if we are to acquire inferential justification for the relevant conclusions. I discuss this view at greater length elsewhere.[12]

Third, there is a strong argument that if there are no a priori probabilities, then there is no knowledge of probabilities whatsoever.[13] To see this, notice first that there is no *foundational* (non-inferential) empirical knowledge of probability. Foundational empirical knowledge consists in direct observations, but no one ever directly *observes* a fact about probability; probability does not look like anything, sound like anything, taste like anything, etc. So knowledge of probabilities will have to be inferential or a priori (or both). Now, suppose that one knows some probability assessment by inference from observations. For instance, let H be the proposition that the probability of the next coin I flip coming up heads is 50 %. Perhaps I have inferred H from a certain series of observations, E. But this inference itself epistemically depends upon a certain assessment of probability: to be justified in making the inference, I must at least have available some source of justification for believing that E supports H.[14] Furthermore, it is clear that E does not *deductively entail* H. In fact, no series of observations ever *entails* any (non-trivial) assessment of probability. So the support that E provides to H must be probabilistic.

Now, let H' be the proposition that E raises the probability of H. I must have some justification for H'. If this justification is itself inferential, then we've embarked on what looks to be an infinite regress. So suppose the justification for H' is non-inferential. We have already said that facts about probability are not directly observed. So H' has non-observational, non-inferential justification. This must be a priori justification. In sum:

1. Some probability assessments are justified.
2. All justification is either (a) non-inferential, a priori justification, (b) justification by direct observation, or (c) inferential justification.
3. No probability assessment is justified by direct observation.
4. All cases of inferential justification for a probability assessment depend upon justification for another probability assessment.
5. There cannot be either an infinite regress or circularity in the chain of justification for a probability assessment.
6. Therefore, not all justification for probability assessments can be inferential (from 4, 5).
7. Therefore, some probability assessments have non-inferential, a priori justification (from 1, 2, 3, 6).

[12] Huemer (2016).

[13] For another argument for this conclusion, based on probability theory, see Huemer (2009, pp. 26–29).

[14] This premise, the "Principle of Inferential Justification," has been defended by several epistemologists, including Fumerton (1995, p. 36) and Huemer (2016).

5.4 Is the principle of indifference inconsistent?

There is a well-known objection to the Principle of Indifference, based upon the fact that there are often different ways of dividing up a set of possibilities. If one applies the Principle of Indifference to different partitions of a given space of possibilities, one can derive inconsistent results. For instance, suppose we know only that S traveled 100 miles in somewhere between 1 and 2 h time. Equivalently, we know that S traveled 100 miles, at a speed between 50 and 100 miles per hour. What is the probability that the trip lasted between 1 and 1.5 h? If we assign a uniform probability density over the possible durations of the journey, we get one result (1/2); but if we assign a uniform probability density over the possible speeds of the journey, we get a different result (2/3).

Very briefly, here is what I think about the problem. First, in the example just given, the correct answer to the problem is 2/3. The reason is that, given a fixed travel distance, the velocity of the traveler *causally explains* the time that the trip takes, rather than the other way around.[15]

The second and more relevant point is that one need not believe that *all* applications of the Principle of Indifference are correct, to believe that *some* applications are correct. In the example of the 100-mile journey, I claim that the application of the Principle of Indifference to the possible *times* is incorrect, while the application to the possible *velocities* is correct. Now, what about the case of the BIVH?

In assessing the probability of one's having experiences like those of a normal person, given that one is a brain in a vat, I considered two ways of applying the Principle of Indifference: one approach is to apply the Principle of Indifference to the possible sequences of experiences that a BIV could be stimulated to have. The other approach is to apply the Principle of Indifference to the possible sets of characteristics of the scientists and their brain-stimulation apparatus.

Which of these two applications is correct? I have not answered this question before now, because it doesn't matter. In fact, I think that the second application is correct; however, *either* application leads to a strong preference for RWH over BIVH, and that is all that matters for purposes of assessing the skeptic's argument.

How can we tell what are the correct applications of the Principle of Indifference? I know no complete answer to this (though I have mentioned one guiding principle, which concerns explanatory priority). But we do not need a complete answer to this. It is possible to identify some F's, even if one does not possess a completely general criterion of F's. For instance, I can identify some cases of knowledge, even though I know of no counter-example-free definition of "knowledge." The applications of the Principle of Indifference proposed in Sects. 3.2 and 3.3 are sufficiently natural, and plausible alternatives are sufficiently elusive, that it is reasonable to presume that one of them is correct. That is, if a skeptic wants to maintain that the correct application of the Principle of Indifference in this case is one that actually results in P(E|BIVH) being comparable to or greater

[15] For explanation, see Huemer (2009, pp. 10–13).

than P(E|RWH), the burden would be on the skeptic to identify this alternative application.

5.5 Of skepticism about probability

I have responded to those philosophers who are skeptical specifically about a priori probabilities. But what if the external world skeptic undertakes to be skeptical about probability assessments in general?

Here, I will just make some brief remarks, taking for granted my view about the relationship between probability and epistemic justification. Suppose one thinks that one can never know, or be justified in believing, any (non-trivial) probability assessment. This view would support a *second-order* skepticism, specifically, the thesis that we *do not know whether* we are justified in external world beliefs. But it would not support first-order skepticism, the thesis that we in fact *are not justified* in our external world beliefs.

Second-order skepticism is not the target of this paper; here, I am interested only in the first-order skeptic. First-order skepticism is an assessment of our degree of justification for external-world beliefs, just as the anti-skeptical position is; the skeptic and the anti-skeptic simply disagree about whether our external world beliefs have a high or a low degree of justification. Thus, both are equally stymied if we cannot assess the degree of justification of these beliefs. An inability to assign relevant probabilities translates into an inability to assess degrees of justification. Thus, the (first-order) skeptic should hope we have some way of assessing probabilities. To put the point another way: if we cannot say what the probability of A is, this obviously does not mean that the probability of A is low; therefore, it does not mean that A is unjustified. Of course, this is predicated on the assumption that claims about justification should be understood (partly) in terms of probability.

6 Extensions: other skeptic-like theories

6.1 An alternative to the theory of gravity

The year is 1787. Kepler's laws of planetary motion are known on the basis of astronomical observation. They state that (1) the orbit of each planet is an ellipse with the sun at one focus, (2) a line joining the planet to the sun sweeps out equal areas in equal times, and (3) the square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit. Isaac Newton has just published his theory of gravity, which proposes an inverse square law of attraction between any two massive bodies.[16] Newton's *Principia* explains mathematically how Kepler's laws of motion can be derived from the law of gravity, together with the three laws of motion that explain the motions of bodies here on Earth. This book

[16] Newton (1999).

will ultimately be considered one of the greatest achievements in the history of science.

But suppose now that a philosopher with a bit of background in epistemology arrives on the scene. "That's all very nice," says the philosopher. "I see how your theory can explain the observed motions of the planets. But I have another possible explanation: maybe God is moving the planets around by direct divine will, and He simply chooses to move them in paths that satisfy Kepler's laws. There is no evidence against my theory, so we are not justified in believing Newton's theory of gravity."

Since this is 1787, no one has any problem with the notion of a God capable of moving physical objects around by direct acts of will. Nevertheless, the philosopher's theory is laughed out of court. Why?

6.2 The problem with god theories

A natural reaction is that the God theory is "too easy." No matter what the motions of the planets had been like, one could have said, "They are like that because God directly wills them to be that way." Indeed, no matter what *anything* had been like, one could have said, "It is that way because God so wills it."

But so what—what is wrong with a theory that provides an easy way of explaining things?

The likelihood account explains our intuitions. The likelihood of the God theory, given the observed motions of the planets, is extremely low, precisely because the God theory can accommodate any set of motions about as well as any other. Motions that approximately satisfy Kepler's laws comprise only a minuscule portion of all of the possible ways of moving; hence, the probability of Keplerian motion would be extremely low.

To appreciate just how low the God theory's likelihood is, notice that, if God is going to be controlling the motions of the planets, there is no necessity for Him to will the planets to follow consistent paths over time. God could have willed the planets to follow Kepler's laws up until January 1, 1787, and then do something else after that. Or He could have willed them to follow Kepler's laws up until January 2, 1787, and then do something else after that. And so on. There is an essentially unlimited range of alternatives. Thus, the probability of our actual observations, given the theory that God directly wills the motions of the planets, is approximately zero.

God theories are available for almost everything else, and the same point applies. Why do I have a flat tire? God willed the tire to flatten. Why did the Soviet Union collapse? God willed it to collapse. Why did this book fall off the shelf? God willed it to fall. All of these are lame theories, precisely because they provide such a universal explanatory strategy. If the tire had stayed inflated, the Soviet Union had persisted, and the book had stayed on the shelf, it could with equal plausibility have been suggested that God willed these things too.

6.3 Other skeptical scenarios

The BIV hypothesis is like the God theories: no matter what happens—no matter whether our experiences are coherent or incoherent, whether objects follow consistent laws or not—the skeptic can always say, "Well, the scientists decided to make it that way."

It is not just the BIV hypothesis that has this flaw. Every skeptical scenario is like that. No matter what happens, one can always explain one's experiences by saying, "Maybe God willed me to experience that," or "Maybe I dreamed that." If one's experiences seem much more vivid and coherent than normal dreams, one can always say, "It's an especially vivid and coherent dream." Again, this flexibility in the skeptical scenarios translates into low likelihoods relative to any given characterization of experience.

Why is the real-world hypothesis not subject to the same criticism—that no matter what happens, one could always say, "That was a veridical perception"? In a strict sense, this may be true; perhaps any series of sensory experiences *could* count as veridical perceptions. It isn't that RWH *entails* coherent experiences; it is just that RWH generates *stronger expectations* about the coherence (and other features) of experience than the skeptical scenarios do. For example, as I consider the (apparent) table at which I am now (seemingly) seated, how should I expect it to behave? If I am a normal person whose experiences come from reliable perceptual faculties, then my background beliefs about how the world works, including such things as tables, are probably by and large correct. I should then expect the table to behave, roughly speaking, like a normal table—for example, to support the weight of a glass, to persist over time, not to rapidly change shape and color, not to bite me, and so on. These expectations are of course not 100 % certain, but they are very probable.

If, however, the world around me is an illusion generated by some deceiver(s), such as God or the brain-manipulating scientists, then there is no need for things to continue to behave in accordance with my background beliefs about how the world works. The regularities in my experience are not generated directly by the laws of nature, but rather by the intentions of the clever deceivers, so any of the regularities could be changed at any moment, if the deceivers so choose.

6.4 A digression on theistic explanations

The above discussion of the weakness of theistic explanations for planetary motion and for sensory experience naturally suggests the further question: are theistic explanations *in general* flawed in the same way, including the sort of theistic explanations that are commonly advanced by actual theists? For instance, it might be said that theism can be used to explain the origin of the universe, the existence of life, the existence and content of morality, the freedom of the will, and/or the laws of nature.

This subject merits lengthier attention than I can give it here. Here I will just make a few brief remarks. First, the way in which God is invoked in popular discourse seems to me to be uncomfortably similar to the brain-in-a-vat style of

explanation. Not, of course, in the sense that theism is used to defend skepticism, but in the sense that God is commonly invoked to provide a kind of explanation that, if legitimate, could be used to explain anything about equally well. For instance, the theistic account of why there are laws of nature is BIV-like in the sense that, no matter what the laws had been like—indeed, even if events around us had been utterly random—one could have said, with no less plausibility, that God made things that way. This flexibility in the God hypothesis prevents the theory from being supported by any evidence.

To sidestep this kind of objection, theists must offer some understanding of the nature of God such that we could make some reasonable inferences regarding what sort of actions God would undertake. For instance, a theist might say that because God is morally good, He would only have created a morally good world; thus, it is not true that God could be invoked to explain just *anything*. I think this a reasonable reply; however, it should be noted that this kind of reply invites attempts by atheists, in turn, to show that the world as we observe it is not in fact the way we would expect a God-designed world to be—thus, for example, the problem of evil will become all the more pressing for the theist.

7 Concluding thoughts

7.1 Summary of the argument

Is there a nonzero probability that I am a brain in a vat? Of course; that's boring. Practically everything has a nonzero probability all the time, and who cares? What is interesting is why the brain in a vat scenario is a bad explanation for sensory experience, and what this case teaches us about the criteria for an adequate explanation.

The BIV hypothesis is one of a family of theories that can be generated to explain any evidence. For instance, one can explain the motions of the planets by citing the hypothesis that God pushes the planets around just so. One can explain why a book fell off a shelf by hypothesizing that an invisible demon pushed it off. What do these theories have in common? They all propose a mechanism that allegedly generated the evidence to be explained, where this mechanism could, with about equal plausibility, be invoked to explain an extremely wide variety of evidence—perhaps even to explain any possible evidence in the particular field in question (for instance, any possible series of planetary motions, or any possible series of sensory experiences). Thus, the likelihood functions for such theories are *spread out* over a wide range of possible evidence. By contrast, the sort of theories we normally take seriously—such as our common sense beliefs about the external world, or Newton's theory of gravity—have relatively *narrow* likelihood functions, that is, they generate substantive expectations about what the evidence should be like (see Fig. 2). If those expectations are realized, the normal theories will receive higher posterior probabilities than the BIV-like theories, in accordance with Bayes' Theorem.

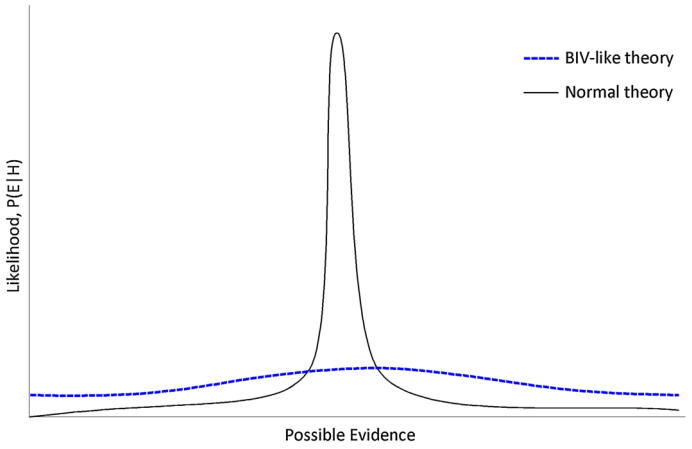


Fig. 2 Normal theories have more concentrated likelihood functions, as compared with BIV-like theories. The area under each *curve* must be 1, since it represents a total probability

Of course, one can generate BIV-like theories that have narrow likelihood functions, by simply incorporating postulates into the theories that stipulate that the mechanism generating the evidence operates in the specific way, or under the specific conditions, that it would need to in order to generate just that evidence and no other. But then one simply trades a theory with a low likelihood for a theory with a low prior probability.

7.2 Serious or skeptical?

What if a skeptic were to come up with a theory that did not have this problem? Suppose the theory postulated some very different mechanism generating our experiences from the one we think generates them, and suppose this mechanism could *not* be used to explain a wide variety of other experiences, but could only be used to explain experiences more or less like ours. Would we then concede that skepticism was correct?

My answer is that we would probably take the theory seriously … but we would not consider it a *skeptical* theory in the sense in which the brain in the vat and the deceiving God theories are skeptical theories. We would simply consider it a new theory about the nature of reality, the way we think about new scientific theories. For example, in the sixteenth century, it was thought that the Sun orbited the Earth, partly because this is the way things look. It was thought that the planets moved around the Earth in more complicated paths (involving epicycles and deferents), because this seemed the most straightforward way of explaining the observed successive positions of the planets in the night sky. Copernicus proposed an alternative theory, on which the Earth rotated on its axis while both the Earth and the planets orbited the Sun.[17] This theory called into question widely accepted

[17] For a brief explanation of the competing accounts of planetary motion, especially retrograde motion, see Smith (2013).

theoretical beliefs, as well as the simple appearance, accessible to everyone, that the Sun moves around the Earth. But Copernicus' theory *was not a skeptical scenario*, and Copernicus was not an "astronomical skeptic," in the sense in which David Hume was an inductive skeptic and John Mackie a moral skeptic.[18] Copernicus' theory was a normal scientific theory. It was not a generic theory applicable to any possible evidence; it was specific to the sort of observations we actually had. For example, the theory that the Earth and Mars are orbiting the Sun at different speeds can be used to explain retrograde motion; it could *not* be invoked to explain just any motion. In fact, if retrograde motion were *not* observed, then we would be under pressure to reject the heliocentric model. That is why the heliocentric model was a serious theory, even before we had very strong evidence for it, and not merely an idle speculation to be entertained in science fiction movies and philosophical thought experiments.

References

Beebe, J. (2009). The abductivist reply to skepticism. *Philosophy and Phenomenological Research, 79*, 605–636.

BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.

Carnap, R. (1962). *Logical foundations of probability* (2nd ed.). Chicago: University of Chicago Press.

Descartes, R. (1984). *Meditations on first philosophy*. In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), *The philosophical writings of Descartes* (Vol. 2). Cambridge: Cambridge University Press.

Dretske, F. (1981). The pragmatic dimension of knowledge. *Philosophical Studies, 40*, 363–378.

Fumerton, R. (1995). *Metaepistemology and skepticism*. Lanham, MD: Rowman & Littlefield.

Fumerton, R. (2004). Epistemic probability. *Philosophical Issues, Epistemology, 14*, 149–164.

Gillies, D. (2000). *Philosophical theories of probability*. London: Routledge.

Huemer, M. (2000). Direct realism and the brain-in-a-vat argument. *Philosophy and Phenomenological Research, 61*, 397–413.

Huemer, M. (2001a). The problem of defeasible justification. *Erkenntnis, 54*, 375–397.

Huemer, M. (2001b). *Skepticism and the Veil of Perception*. Lanham, MD: Rowman & Littlefield.

Huemer, M. (2009). Explanationist aid for the theory of inductive logic. *British Journal for the Philosophy of Science, 60*, 1–31.

Huemer, M. (2016). Inferential appearances. In B. Coppenger & M. Bergmann (Eds.), *Intellectual assurance: Essays on traditional epistemic internalism.* Oxford: Oxford University Press (forthcoming).

Hume, D. (1975). *Enquiry concerning human understanding*. In L. A. Selby-Bigge & P. H. Nidditch (Eds.), *Enquiries concerning human understanding and concerning the principles of morals* (3rd ed.). Oxford: Clarendon Press. **(Originally published 1748)**.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge: Cambridge University Press.

Keynes, J. (1921). *A treatise on probability*. London: Macmillan.

Klein, P. (1995). Skepticism and closure: Why the evil genius argument fails. *Philosophical Topics, 23*, 213–236.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York: Penguin.

Newton, I. (1999). *The principia: Mathematical principles of natural philosophy*. Berkeley, CA: University of California Press. **(Originally published 1787)**.

Pollock, J., & Cruz, J. (1999). *Contemporary theories of knowledge*. Lanham, MD: Rowman & Littlefield.

Putnam, H. (1981). Brains in a vat. In *Reason, truth, and history* (pp. 1–21). Cambridge: Cambridge University Press.

[18] Hume (1975), section IV; Mackie (1977).

Ramsey, F. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (pp. 156–198). London: Kegan, Paul, Trench, Trubner & Co.

Smith, S. (2013). *Retrograde motion*. http://www.lasalle.edu/∼smithsc/Astronomy/retrograd.html. Accessed October 10, 2013.

Vogel, J. (1990). Cartesian skepticism and inference to the best explanation. *Journal of Philosophy, 87*, 658–666.

Vogel, J. (2010). BonJour on explanation and skepticism. *Studies in History and Philosophy of Science, 41*, 413–421.