A-Symmetric Confirmation and Anthropic Skepticism

Benjamin Eva*

July 9, 2017

Abstract

In recent years, anthropic reasoning has been used to justify a number of controversial skeptical hypotheses (both scientific and philosophical). In this paper, we consider two prominent examples, viz. Bostrom's 'Simulation Argument' and the problem of 'Boltzmann Brains' in big bang cosmology. We argue that these cases call into question the assumption, central to Bayesian confirmation theory, that the relation of evidential confirmation is universally symmetric. We go on to argue that the fact that these arguments appear to contradict this fundamental assumption should not be taken as an immediate refutation, but should rather be seen as indicative of the peculiar role that the relevant hypotheses play in their respective epistemic frameworks.

1 Introduction

Let H be some hypothesis, and let E be a prospective piece of evidence for believing H. In Bayesian confirmation theory, the degree to which E confirms H is given in terms of the conditional probability $P(H|E) = \frac{P(E \wedge H)}{P(E)}$. Specifically, E confirms H if and only if $P(H|E) > P(H)$, i.e. if the probability of H being the case, given that E is the case, is higher than the prior probability of H being the case.

*Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – http://be0367.wixsite.com/benevaphilosophy – benedgareva@icloud.com.

Now, we call the evidence-hypothesis pair (E, H) 'confirmationally a-symmetric' (or 'a-symmetric' for short) if and only if E confirms H but H *disconfirms* E (equivalently, if H confirms $\neg E$). In such a case, E is evidence for believing H, but H is also evidence for believing $\neg E$. It is not hard to see that the Bayesian definition of confirmation precludes the existence of a-symmetric evidence-hypothesis pairs, since $P(H|E) > P(H)$ implies that

$$P(E|H) = \frac{P(E \wedge H)}{P(H)} > \frac{P(E \wedge H)}{P(H|E)} = \frac{P(E \wedge H)}{\frac{P(E \wedge H)}{P(E)}} = P(E)$$

Our main contention is that this restriction may not be universally justified. In section 2 we consider two important examples of skeptical hypotheses arising from anthropic reasoning in contemporary philosophy and cosmology and argue that the arguments supporting these hypotheses appear to imply the possibility of a-symmetric evidence-hypothesis pairs (henceforth 'EHP's).

We argue that the self-undermining character of these hypotheses should not be interpreted as paradoxical, but rather as indicative of the peculiar role they play in their respective epistemological frameworks. Section 3 considers a possible Bayesian analysis of the prospective a-symmetric EHP's, but finds it lacking. Section 4 considers the possibility of providing a non-Bayesian formalisation of the notion of confirmation that allows for the existence of a-symmetric EHP's, and connects these considerations to Norton's 'material theory of induction'. Section 5 concludes.

2 A-Symmetric EHP's

2.1 The Simulation Argument

Let **SH**[1] be the alarming skeptical hypothesis that 'we are living in a computer simulation'. Since Bostrom (2003) forwarded the now famous 'simulation argument' (henceforth **SA**), **SH** has been the focus of sustained philosophical (see e.g. Weatherson 2003, Lewis 2013) and scientific (see e.g. Barrow 2007) attention. We contend

[1]For 'simulation hypothesis'.

that **SH** belongs to an a-symmetric EHP, in the sense outlined in the previous section. In order to substantiate this claim, we need to look closely at **SA**.[2] First, we introduce some new terminology. By 'posthuman civilization', we will mean a civilization that 'has acquired most of the technological capabilities that one can currently show to be consistent with physical laws and with material and energy constraints' (Bostrom 2003: 245), and by 'ancestor simulation' we will mean a detailed computer simulation of the evolutionary history of the relevant civilization. Then, **SA** has the following form,

Premise 1: The property of consciousness is substrate independent[3]

Premise 2: 'Posthuman civilizations would have enough processing power to run hugely many ancestor simulations' (Bostrom 2003: 248)

Conclusion: Either (1) The human species is very likely to go extinct before it reaches the 'posthuman' stage, (2) Any posthuman civilization is extremely unlikely to run a significant number of ancestor simulations of its own evolutionary history, or (3) It is extremely likely that we are living in a computer simulation.

In what follows, we assume that premise 1 of **SA** has probability arbitrarily close to 1, i.e. we assume the substrate independence of the property of 'consciousness' as a fundamental piece of background knowledge. Secondly, we assume that premise 2 has high but non-maximal probability, i.e. we are not completely certain in our current estimation of the physical limits of computational power, but we are pretty confident.[4] Finally, we take it that, given these assumptions, advocates of **SA** will be happy to accept that the following simplified simulation argument (**SSA**) is a

[2]It should be noted that we will remain entirely agnostic about the soundness of **SA**. Of course, this section will only be of interest to those that view **SA** as a remotely plausible argument. For detailed discussions concerning the soundness of **SA**, see e.g. Weatherson (2003) and Lewis (2013).

[3]i.e. 'It is not an essential property of consciousness that it is implemented on carbon-based biological neural networks inside a cranium: silicon based-processors inside a computer could in principle do the trick as well.' (Bostrom 2003: 244)

[4]These assumptions actually do very little work in what follows, and we could do without them. However, they serve to simplify the analysis significantly, and they don't seem unreasonable in the context of the current dialectic.

good one, in the sense that its premises provide significant evidential confirmation of its conclusion,

Premise 1: The property of consciousness is substrate independent

Premise 2: Posthuman civilizations would have enough processing power to run hugely many ancestor simulations

Conclusion: We are living in a computer simulation.

More specifically, we take it to be uncontroversial that supporters of **SA** should view premise 2 as a significantly confirmatory piece of evidence for **SH**. Learning that premise 2 holds with certainty should increase our degree of belief in **SH**. Thus, according to Bayesian confirmation theory, **SH** should also be good evidence for premise 2. But now, let's suppose we learn that **SH** is the case. What effect would this have for our belief in premise 2? Well, premise 2 is a contingent empirical proposition, and the high degree of belief that we had in it prior to learning **SH** was justified by a huge body of observations about the external world. However, after we learn **SH**, these observations should surely carry less weight than before. For, we now know that they were observations, not of the real world, but of some complex computer simulation of the real world. So, any errors in that simulation of the real world could lead to errors in our beliefs about what is physically possible in the real world, and this includes our best estimations of the maximal physically possible level of processing power.

To make things slightly more precise, first assume that **SH** is true. Then we are living in a computer simulation. There is nothing in **SA** that guarantees that this simulation will be perfect. In fact, we have good reason (see e.g. Barrow 2007) to believe that the simulation will always be subject to some degree of error in its representation of the fundamental physics of the real world. Indeed, it might well be that we are living in a simulation that has some error e in its approximation of a physical parameter λ that is relevant to determining the physical limits of computation. So our estimation of the computing power available to posthuman civilizations will have some error $e' > 0$ proportional to e. Since we know that

such a situation is possible, we should be less confident in our approximations of the processing power available to posthuman civilizations than we were prior to assuming that **SH** holds. Thus, we should decrease our degree of belief in premise 2. So the EHP that consists of premise 2 and **SH** is a-symmetric.[5]

Note that we assume nothing about the scale of the a-symmetry here. It might be that while premise 2 offers very significant evidential support for **SH**, **SH** only undermines premise 2 to a negligible degree. This will generally be the case if we assume that the simulations are always very good. But this doesn't affect our arguments. We only need to show that **SH** disconfirms premise 2. The extent of this disconfirmation is irrelevant.

At this point, one might be tempted to reject **SH** as paradoxical, on the grounds that it undermines its own evidential basis. But this seems *ad-hoc*. For, we have an argument in favour of **SH** and, we argue, **SH** should stand or fall according to the success of the argument. If one is unable to identify any problematic steps in **SA**, then one is bound to accept its conclusion, regardless of whether it undermines its own evidence.

The Bayesian and the critic of **SA** are likely to be unconvinced here. A natural response is something like the following. The premises of **SA**, although independently plausible, lead to a deeply problematic epistemic tension, crystalised in the apparent a-symmetry of the EHP composed of **SH** and premise 2. This tells us 'what is really wrong' with **SA** and gives us reason to reject the whole argument outright. Specifically, premise 2 is problematic because if it is true, then it is probably false. So we should reject premise 2, thereby undermining **SA**.

But it is far from clear that this is the right response here. For, premise 2 is a purely empirical proposition that is justified by our current best scientific understanding of computation and the physical world. We cannot simply reject such a premise without offering either an elucidation of why the premise is not actually justified by current science, or an alternative equally plausible scientific account of the relevant phenomena that does not allow us to infer anything like premise 2. Even

[5]To be clear, I take premise 2 to be evidence for **SH** in the sense that conditioning on premise 2 should raise the probability of SH.

critics of **SA** have tended to accept premise 2 as well motivated, given the sheer weight of empirical evidence on which it is based. One cannot simply reject it as if it were some controversial and abstract philosophical claim whose veracity could only ever be tested by a-priori arguments. In rejecting premise two, we would be denying an *empirical* prediction grounded in current-day scientific knowledge. At a purely methodological level, there is something problematic about such a strategy.

More generally, we contend that the a-symmetric character of the EHP (**SH**, premise 2) is simply a symptom of the unusual epistemological status of the simulation hypothesis, and so should not really come as a surprise. If **SH** is true, then many of the basic assumptions that underpin the epistemological mechanisms we use to find out about the world are founded on false assumptions, and so it is natural to expect that the formal tools we normally use to reason about the world might be insufficient for dealing with such a hypothesis.

2.2 Boltzmann Brains

The second example of an a-symmetric EHP comes from contemporary cosmology. For ease of exposition and accessibility, we omit many of the physical details (the interested reader can see e.g, Boddy and Carroll 2013, Nomura 2015). The main idea is this.

One of the best confirmed models of modern cosmology, the ΛCDM (cold dark matter) model appears to allow for the universe to enter a stage in its evolution (the 'De Sitter vacuum phase') where there will be arbitrarily high numbers of random fluctuations which will eventually 'reproduce any local region of the current universe to arbitrary precision' (Boddy and Carroll 2013: 1), including, for example, intelligent observers. But then, for each time that these fluctuations give rise to a whole coherent universe in which entire species of intelligent observers evolve deterministically, there will be countless numbers of isolated 'Boltzmann brains' (human brains that spontaneously spring into existence complete with randomly generated memories, opinions, perceptions etc) passing in and out of existence. A natural way to think about this is to imagine the universe as being like a 'box of

atoms' moving around over an indefinite period of time. Carroll notes that, in this situation, the atoms will almost always

> ...be in a disordered, high-entropy, equilibrium state. Eventually, just by chance, they will take the form of a smiley face, or Michelangelo's David, or absolutely any configuration that is compatible with what's inside the box. If you wait long enough, and your box is sufficiently large, you will get a person, a planet, a galaxy, the whole universe as we now know it. But given that some of the atoms fall into a familiar-looking arrangement, we still expect the rest of the atoms to be completely random. Just because you find a copy of the Mona Lisa, in other words, doesn't mean that it was actually painted by Leonardo or anyone else; with overwhelming probability it simply coalesced gradually out of random motions. Just because you see what looks like a photograph, there's no reason to believe it was preceded by an actual event that the photo purports to represent. If the random motions of the atoms create a person with firm memories of the past, all of those memories are overwhelmingly likely to be false. (Carroll 2013)

The upshot of all this is that, according to an extraordinarily well confirmed and important model from contemporary cosmology, it is far more likely that I, together with all of my mental states, am the result of some random physical fluctuation than it is that those mental states were produced by a long-running interaction between a reasonably rational agent and the world around them. I have memories, beliefs and impressions that appear to pertain to the real world, but it is overwhelmingly likely that they are mere byproducts of some stochastic physical fluctuations, i.e. it is overwhelmingly likely, according to a significant body of cosmological evidence, that I am a Boltzmann brain. If we let **E** represent the relevant body of evidence,[6] and let **BB** be the Boltzmann brain hypothesis, 'I am a Boltzmann brain', then we have that $P(\mathbf{BB}|\mathbf{E}) > P(\mathbf{BB})$, i.e. **E** confirms **BB**.

[6]To be clear, it is not important to our arguments that **E** should represent the entire relevant body of cosmological evidence. It is sufficient for our purposes that **E** be any piece of observational cosmological evidence confirming the ΛCDM model.

Now, as before, we can consider what happens when we assume the truth of the skeptical hypothesis. If **BB** is true, then all of my perceptions, memories and experience of the external world, and hence all of my beliefs about cosmology, are randomly generated and completely untrustworthy. I happen to believe in some assortment of the best theories of contemporary cosmology, but I could just as easily have sprung into existence as an advocate of Aristotelian cosmology, or any theory whatsoever. This being the case, I should radically decrease my confidence in all of my beliefs pertaining to cosmology (and pretty much everything else) and adopt the least informative possible probability distribution as my credal state. But this would entail significantly decreasing my degree of belief in **E**, since **E** is just a subset of my beliefs about cosmology. So **BB** disconfirms **E**, i.e. $P(\mathbf{E}|\mathbf{BB}) < P(\mathbf{E})$. So (**E**, **BB**) is a-symmetric.

Before going further, it is worth pre-empting some intuitive responses to the argument outlined above.

Firstly, it is natural to object at this point that the ΛCDM model is *just a model*, and that as such, it is bound to contain a high-number of approximations, idealisations and simplifications, all of which can lead to absurd consequences. Surely then, we can contend that the argument for the existence of Boltzmann brains is one such absurd consequence arising from a pragmatically justified idealisation in the ΛCDM model. If this were the case, then we wouldn't have to take these kinds of considerations seriously in the first case.

Unfortunately, this response is unsatisfactory. The ΛCDM model is extremely well confirmed by vast quantities of cosmological data, and it has an excellent record of predicting new phenomena[7]. Most cosmologists believe that it represents the best currently available picture of what the universe is like on large scales. A crucial feature of the ΛCDM model is the fact that it describes a universe that includes a vacuum energy density Λ[8]. And the argument from the ΛCDM model to the existence of Boltzmann brains relies primarily on (i) The presence of the parameter

[7]For example, the baryon acoustic oscillation feature and the polarisation of the cosmic microwave background.

[8]Λ is also referred to as the 'cosmological constant' or 'dark energy'.

Λ, and (ii) The fact that the universe is expanding at an accelerating rate. These are central features of the ΛCDM model, and they are both necessary for its empirical adequacy. So **BB** doesn't emerge as the result of some unrealistic idealisations in the model. Rather, it is a direct consequence of the model's characteristic physical content. So, insofar as the model is empirically successful and important to the scientific community, we are forced to take **BB** seriously.

And it turns out that the possibility of Boltzmann brains has actually been taken very seriously by parts of the scientific community. To give just one example, De Simone et al (2010) actually attempt to impose bounds on divergent spacetime volume in inflationary multiverse cosmology by defining a measure for the ratio of Boltzmann brains to 'normal observers'. The so called 'problem of Boltzmann brains' has guided a significant amount of research in theoretical cosmology and is clearly viewed in many quarters as being an issue of genuine physical significance, not just a side effect of the way cosmologists happen to construct their models. The importance of the problem is most clearly evinced by the wide range of sophisticated solutions that have been proposed by prominent cosmologists. For example, while Page (2006) explicitly countenances the idea that the **BB** problem might be sufficiently dire to justify the rejection of the cosmological constant model (and to suggest that the De Sitter vacuum phase is not eternal), Carlip (2007) notes that the **BB** problem can be elegantly solved by Steinhardt and Turok's (2002) cyclical model of the universe. Carlip (2007) also considers the possibility of solving the **BB** problem by positing a slow variation in the values of the fundamental constants. Yet another possible solution is offered by Linde (2007) who suggests that inflationary cosmology offers a way out of the **BB** problem (because of the prevalence of younger bubble universes containing more 'normal' observers compared to older bubble universes containing more **BB**'s). Further possible solutions have been suggested, for example, by Gott (2007) and Vilenkin (2006). The crucial point to note is that this is a problem whose solution is by no means considered trivial by the scientific community. There is still no consensus and the problem is taken so seriously that it has genuinely affected the evaluation of prospective cosmological models.

As with the simulation hypothesis, we do not take the apparent a-symmetry of its relation to the supporting evidence to constitute a refutation of **BB**. Rather, we argue that this is another example of an important and philosophically interesting hypothesis that happens to be part of an a-symmetric EHP. Any account of confirmation that rules such hypotheses out simply in virtue of this feature is unsatisfactory. For, as Carroll notes,

> Discussions of the Boltzmann Brain problem typically occur in the context of speculative ideas like eternal inflation and the multiverse (not that there's anything wrong with that). And, let's admit it, the very idea of orderly configurations of matter spontaneously fluctuating out of chaos sounds a bit loopy, as critics have noted. But everything I've just said is based on physics we think we understand: quantum field theory, general relativity, and the cosmological constant. This is the real world, baby. Of course it's possible that we are making some subtle mistake about how quantum field theory works, but that is more speculative than taking the straightforward prediction seriously. (Carroll 2013)

By ruling out[9] **BB** purely on the grounds that it undermines its own evidence, we would be placing ad-hoc and a-priori constraints on the implications of physics, which is surely an unpalatable consequence. Thus, as with the simulation hypothesis, we should be willing to simply treat the a-symmetric relationship of **BB** to its evidence as a new, admittedly unintuitive, *feature* of its place in our epistemic framework.

One important point to note here is that, while the scale of the a-symmetry between **SH** and its evidence will generally be very modest, this is not the case for **BB**. It seems clear that the a-symmetric relationship between **BB** and the relevant cosmological evidence is quite extreme. This follows from the fact that, while **SA** ensures that our confidence in our beliefs after conditioning on **SH** will still be

[9]To clarify, of course Bayesian confirmation theory does not itself rule out the consideration of **BB**. It simply precludes the possibility of a-symmetric EHP's. But since there appear to be good reasons for claiming that **BB** is part of an a-symmetric EHP, advocates of Bayesian confirmation theory may be tempted to disregard the hypothesis as absurd.

quite high (since we are guaranteed that the ancestor simulations have a pretty-high level of fidelity), no such guarantee is given in the **BB** case. So the degree of the a-symmetry between a hypothesis and its evidence can, it seems, vary in quite a radical way.

Again, we argue that the a-symmetric relationship between **BB** and the relevant cosmological evidence is a natural expression of the skeptical nature of the hypothesis. Unlike traditional Cartesian style skeptical hypotheses, **BB** and **SH** are supported by large bodies of contingent empirical evidence. However, once we consider the implications of this evidence and accept the associated skeptical hypotheses, the epistemological procedures that justified our prior belief in that evidence are undermined, and we are forced to decrease our confidence in it.

Before moving on, one more clarification is required. It might be thought that the very idea of an a-symmetric EHP is *inconsistent* and leads immediately to contradiction. The argument might go something like this. If (E, H) is an a-symmetric EHP such that (i) the prior probability of E is high, (ii) the conditional probability $P(H|E)$ is high, and (iii) the conditional probability $P(E|H)$ is low[10], then it is clear that we have good reason for believing both H and E, and we *should* have a high credence in them. But this can't be right! For, if H is the case then I should have a low degree of belief in E, so how can I can possibly believe them both to a high degree? It looks like I both should and should not have a high degree of belief in both H and E.

But this argument relies crucially on an informal statement of Bayes' theorem. In order to make the argument precise we will, of course, have to use Bayes' theorem to derive the conditional probability of E given H, which will indeed lead straight to a contradiction. So, assuming a Bayesian framework, the idea of a-symmetric EHP's is unsurprisingly inconsistent. Given that we started from the observation that Bayesian confirmation theory precludes the possibility of any a-symmetric EHP's, this is not very interesting. So the Bayesian cannot simply dismiss the challenge of a-symmetric confirmation as inconsistent, since the proof of inconsistency itself

[10]The Boltzmann brains example satisfies these three conditions.

relies on Bayesian tools.[11]

In section 4 we will see that there already exists a robust and coherent form of inductive logic that is able to accommodate a-symmetric EHP's without surrendering consistency. This being the case, the fact that EHP's are inconsistent within a Bayesian setting carries little weight.

3 A Bayesian Response

A Bayesian might respond to these arguments by offering something like the following simple model for **BB**[12]. Let H_1 and H_2 be two theories of physics that, for current purposes, we take to be mutually exclusive and jointly exhaustive, and suppose that while H_1 implies the existence of many Boltzmann brains for each actual human being, H_2 precludes the existence of Boltzmann brains. Suppose further that E_1 and E_2 are pieces of confirmatory evidence for H_1 and H_2, respectively. To make things clear, let's fix the priors as follows,

- H_1 implies that there are 100 Boltzmann brains, which randomly observe either E_1 or E_2, and 1 'real' observer who will of course have to observe E_1.
- H_2 implies that there are no Boltzmann brains, and 1 'real' observer who will of course have to observe E_2.

[11] Here, one might be tempted to reply that Bayesianism is an extremely well supported and powerful epistemological framework. Thus, since the notion of a-symmetric confirmation is inconsistent within a Bayesian setting, the sheer weight of pragmatic and theoretical considerations on the side of Bayesian confirmation theory should be sufficient for us to dispel the idea. However, this kind of reasoning is at best overly conservative and at worst downright dogmatic. Of course, I don't deny the manifold virtues of Bayesian epistemology, but the fact that Bayesianism is well supported and useful does not constitute a principled justification for ignoring those aspects of scientific reasoning that are incompatible with Bayesian principles like the universal symmetry of confirmation. The literature is replete with examples of cases where the Bayesian is unable to give a satisfactory account of manifestly rational forms of inductive reasoning (see e.g. Norton (2010)). For example, it is well known that the principle of indifference, which states that a rational agent who is indifferent over several possible outcomes should not assign any outcome a higher degree of belief than any other, leads immediately to inconsistency and paradox when formulated in a Bayesian setting (see e.g. Norton (2010), Rinard (2013), Van Fraassen (1989)). It would seem ad-hoc and dogmatic to simply reject the principle of indifference because it is inconsistent with classical Bayesianism. And indeed, several authors have suggested amending the standard Bayesian framework in order to resolve the paradoxes of indifference (see e.g. Joyce (2005), Weatherson (2007), Norton (2010)). This is a more fruitful and principled response. The mere fact that Bayesianism has achieved many successes does not mean that we should ignore its failures.

[12] A similar analysis could be given for **SH**.

- H_1 and H_2 have equal prior probabilities $P_0(H_1) = 1/2 = P_0(H_2)$.

This setup implies that $P_0(\mathbf{BB}) = \frac{100}{102}$. But suppose that we now observe E_1, and update to the posterior distribution $P_1 = P_0(-|E_1)$. By definition, E_1 confirms H_1 and disconfirms H_2, so $P_1(H_1) > P_0(H_1)$. Similarly, the probability of **BB** will obviously increase, since it has high probability for H_1 and zero probability for H_2. So E_1 confirms both **BB** and H_1. Conversely, if I assume **BB**, then this rules out H_2 and so decreases the probability of E_2 being observed, thereby increasing the probability of E_1. So **BB** confirms E_1 and vice-versa. There is no a-symmetry. So, claims the Bayesian, **BB** is not problematic for the Bayesian approach to confirmation.

But this kind of response is question begging. For, the Bayesian model offered above simply assumes that **BB** confirms E_1, which is precisely the point that is being disputed. There appear to be compelling reasons to claim that the Boltzmann Brains hypothesis undermines its own evidence. The fact that the Bayesian model offered above violates this property simply tells us that the model is not describing the kind of situation in which we are interested. The temptation to simply 'bake in' the Bayesian assumption that confirmation is always symmetric is a strong one, but it leaves us unable to properly model interesting cases like **BB** and **SH** and, as such, it should be resisted.

More generally, we contend that the only philosophically satisfactory approach to precluding the possibility of a-symmetric EHP's would be to show that the purported examples are not actually supported by the relevant items of evidence. Both **BB** and **SH** are supported by a controversial form of anthropic reasoning that involves notions of typicality and assumes that uninformative priors should always be represented by flat distributions. Thus, critics (see e.g. Norton 2010) of anthropic reasoning will probably not be greatly concerned by the examples given here. However, there are many (see e.g. Bostrom 2003, Armstrong 2011) who defend the use of anthropic reasoning in philosophy and science. Indeed, anthropic reasoning plays a very important role in contemporary cosmology (see e.g. Weinberg, 1987). So identifying the faults in the relevant arguments will not be a trivial matter, and in the absence of a comprehensive refutation of the anthropic reasoning that supports **BB**

and **SH**, we are compelled to take seriously the possibility of a-symmetric EHP's.

The Bayesian might also respond by claiming that Bayes' theorem tells us how we *ought* to reason, i.e. that it *should* never be the case that a piece of evidence E increases our degree of belief in an hypothesis H, while H decreases our degree of belief in E. To reason in this way is to be epistemically inconsistent, and while it might be possible for actual epistemic agents to be confused in this way, an *ideal* epistemic agent never would be. Bayesian confirmation theory, according to this response, is primarily a normative theory, and the possibility of agents being in a situation where they happen to judge the relationship between some hypothesis and some evidence as a-symmetric has no bearing on the validity of Bayesian norms.

However, there is a problem with this response. Specifically, in the two cases considered here, it seems that it is *correct* to judge the relationship between the elements of the relevant EHP's as being a-symmetric. If I believed that **BB** was the case, then I really should be less certain in my beliefs about the cosmological evidence, but that same cosmological evidence really should make me more confident that **BB** is true. We are not just arguing that some confused agent happens to be in a belief state that includes a-symmetric EHP's. We are arguing that the current epistemic situation licenses and indeed requires us to regard the confirmatory relationship between the items of the relevant evidence-hypothesis pairs as a-symmetric. So simply emphasizing the normative interpretation of Bayesian confirmation theory won't be sufficient.

4 Non-Bayesian Confirmation

In an influential series of papers (see e.g. Norton 1994, 2007, 2010, forthcoming) John Norton has developed and defended a new approach to inductive inference. Against the Bayesian paradigm, Norton has disputed the existence of any one privileged and universally applicable form of inductive logic. Thus, according to Norton, one's choice of inductive logic should always be influenced to some extent by the material facts of one's inductive situation. This observation grounds what he calls the 'material theory of induction'. Pertinently for our purposes, Norton has devel-

oped a general procedure for defining new forms of inductive logic, the properties of which can vary greatly depending on one's requirements. Broadly speaking, the procedure is as follows.

First, let B represent the complete, bounded and atomic Boolean algebra of equivalence classes of logically equivalent sentences in the relevant language L (also known as the 'Lindenbaum algebra' of L). An inductive logic is then a way of assigning, to any elements $a, b \in B$ an 'inductive strength' $[a|b]$, representing the extent to which a is inductively supported by b. In order to justify this interpretation of the value $[a|b]$, we need to ensure that different inductive strengths can be compared in a meaningful way. In order to do so, we assume that there exists a partial ordering on the set of inductive strengths, So $[a|b] < [a|c]$ means that a derives more inductive support from c than it does from b. In order to allow for the existence of non-trivial limits, Norton also requires that the partial ordering on inductive strengths is dense, i.e. for any $x, y \in B$ with $x < y$, there exists some $z \in B$ such that $x < z < y$ holds. Finally, it is also assumed that the partial ordering has both a unique maximum and a unique minimum value. We require that this maximum value is assigned whenever b deductively entails a (equivalently, when $b < a$ holds in B), and the minimum value is obtained whenever b entails $\neg a$ (equivalently, when $b < \neg a$ holds in B).

Now, in order to obtain a particular instance of such an inductive logic, we of course need a rule for assigning inductive strengths to ordered pairs of elements from B. Norton (forthcoming) focuses especially on the class of 'deductively definable' inductive logics, i.e. those for which the inductive strength assigned to an ordered pair of elements of B is a function purely of the deductive relationship between those elements (as encoded by the ordering relation on B). To clarify the constraint of deductive definability, consider the simple example where $X = x_1$, $Y = x_1 \vee x_2 \vee x_3$, $X' = x_2$ and $Y' = Y = x_1 \vee x_2 \vee x_3$, for some logical atoms x_1, x_2, x_3. In this case, deductive definability requires that

$$[X|Y] = [x_1|x_1 \vee x_2 \vee x_3] = [X'|Y'] = [x_2|x_1 \vee x_2 \vee x_3]$$

This is because the deductive relationship between x_1 and $x_1 \vee x_2 \vee x_3$ is the same as the deductive relationship between x_2 and $x_1 \vee x_2 \vee x_3$. In both cases, the element

on the left is entailed by the disjunction on the right (of which it is a constituent disjunct). So, by the requirement that inductive strength should be a function of the deductive relation between the relevant elements, we obtain the restriction that $[X|Y] = [X'|Y']$ in any deductively definable inductive logic.

But even when we restrict our attention to deductively definable logics, there are still many possibilities. For example, if we want to obtain the standard inductive logic of Bayesian confirmation theory, we can use the following definition, where $N(a)$ denotes the number of atoms in B that deductively entail a.

$$[a|b] = \frac{N(a \wedge b)}{N(b)}$$

We can also represent the prior probability of a by $[a|K]$, where K denotes our background knowledge. It is easy to see that with these definitions, the deductively definable inductive logic corresponding to standard Bayesian confirmation theory rules out the existence of a-symmetric EHP's. However, there are other forms of deductively definable inductive logic that do not impose this restriction. In particular, Norton's 'specific conditioning' logic (see section 11.2 of Norton (2010)), defined below, allows for the possibility of a-symmetric EHP's[13].

$$[a|b] = \frac{N(a \wedge b)}{N(b)} \cdot \frac{N(a \wedge b)}{N(a)}$$

To see that this choice of inductive logic allows for a-symmetric EHP's, note first that this definition implies the surprising equality $[a|b] = [b|a]$. Next, suppose that we have the following setup, where K represents our background knowledge,

- $N(K) = 30$, $N(a \wedge b) = 2$, $N(a \wedge \neg b) = 1$, $N(\neg a \wedge b) = 7$, $N(\neg a \wedge \neg b) = 20$

Then we have the following prior probabilities for a and b,

- $[a|K] = 3/30 = 0.1$, $[b|K] = 9/30 = 0.3$

Together with the following likelihoods,

- $[a|b] = [b|a] = 2/9 \cdot 2/3 = 0.148$

[13]Many thanks to John Norton for pointing this out to me with and providing the following example.

Putting this all together, we have that $[a|b] = 0.148 > [a|K] = 0.1$ meaning that b confirms a, but $[b|a] = 0.148 < [b|k] = 0.3$, meaning that a disconfirms b. So a and b are an a-symmetric EHP in the deductively definable specific conditioning inductive logic.

Now, the moral of this story isn't that specific conditioning is necessarily the right inductive logic to use in the kinds of cases considered in this paper. The point is simply to show that there are perfectly rigorous and coherent non-Bayesian forms of inductive logic in which a-symmetric EHP's can be consistently and meaningfully formalised. There is nothing inherently contradictory or paradoxical about the idea of an a-symmetric EHP, once one moves beyond Bayesian confirmation theory. And insofar as this paper has made a convincing case for the claim that contemporary science does indeed require us to seriously countenance this kind of epistemological phenomenon, we have good reason to think that these other inductive logics might have an important role to play.

5 Conclusion

We've presented two prospective examples of a-symmetric EHP's that, we argue, cannot be properly understood in the confines of traditional Bayesian confirmation theory. Furthermore, we've argued that these examples, far from being idle Cartesian style skeptical hypotheses, are based on our current best knowledge of the physical world. Both **SH** and **BB** are taken seriously in the relevant scientific communities, and for philosophers to simply label them as paradoxical would be a failure of the discipline.

We remain agnostic about the arguments that have been forwarded in support of **BB** and **SH**, but insofar as we are unable to provide a comprehensive refutation of these arguments (which, given the lack of consensus in the literature, seems to be the case), we need to seriously consider the task of accommodating within our epistemology hypotheses that undermine their own evidence. Section 4 hinted at one possible way of achieving this.

Of course, as we noted earlier, the considerations discussed here will only be

interesting to those without principled objections to the kinds of anthropic reasoning used in the relevant arguments. We take no side about whether or not these are indeed good arguments, but we do contend that rejecting the arguments purely in light of their implications regarding a-symmetric EHP's is not a philosophically legitimate maneuver. However, we are open to the possibility that there is something else wrong with the arguments, and that the confirmational a-symmetry might be a consequence of such a flaw. If that were the case, there would indeed be no reason to worry about these issues. But there is simply no consensus on whether the arguments are flawed (and if so, in what way).

It should be noted that we are not advocating any kind of general rejection of the Bayesian approach to confirmation. Indeed, we accept that this is the best general account of confirmation that is currently available to us. However, the arguments presented here suggest that, for a special class of skeptical hypothesis, the Bayesian approach to confirmation breaks down. Given the current prevalence of such hypotheses in philosophy and beyond, this is an important limitation that needs to be addressed.

Acknowledgements

This work was generously supported by the Ludwig Maximilian University Center for Advanced Studies. I'd also like to thank John Norton and three anonymous referees for their helpful comments on earlier versions of the paper.

References

[1] Armstrong, S. (2011). Anthropic Decision Theory for Self Locating Beliefs. *Journal of Philosophy*. https://arxiv.org/abs/1110.6437

[2] Barrow, J. (2007). Living in a Simulated Universe. In *Universe or Multiverse*: 481-486. Cambridge: Cambridge University Press

[3] Boddy, K. and Carroll, S. (2013). Can the Higgs Boson Save Us From the Menace of the Boltzmann Brains? https://arxiv.org/abs/1308.4686

[4] Bostrom, N. (2003). Are You Living In a Computer Simulation? *Philosophical Quarterly* 53(211): 243-255

[5] Breuckner, A. (2008). The Simulation Argument Again. *Analysis* 68(3): 224-226

[6] Carlip, S. (2007) . Transient Observers and Variable Constants or 'Repelling the Invasion of the Boltzmann Brains'. *Journal of Cosmology and Astroparticle Physics*. DOI: https://doi.org/10.1088/1475-7516/2007/06/001

[7] Carroll, S. (2013). The Higgs Boson vs. Boltzmann Brains. http://www.preposterousuniverse.com/blog/2013/08/22/the-higgs-boson-vs-boltzmann-brains/

[8] De Simone, A., Guth, A., Linde, A., Noorbala, N., Salem, M. and Vilenkin A. (2010). Boltzmann brains and the scale-factor cutoff measure of the multiverse. *Physical Review D* 82: 063520

[9] Gott, J. (2008). Boltzmann Brains: I'd rather see than be one. https://arxiv.org/abs/0802.0233

[10] Joyce, J. (2005). How Probabilities Reflect Evidence. *Philosophical Perspectives* 19(1): 153-178

[11] Lewis, P. (2013). The Doomsday Argument and The Simulation Argument. *Synthese* 190(18): 4009-4022

[12] Linde, A. (2007). Sinks in the landscape, Boltzmann brains and the cosmological constant problem. *Journal of Cosmology and Astroparticle Physics*. DOI: https://doi.org/10.1088/1475-7516/2007/01/022

[13] Nomura, Y. (2015). A Note on Boltzmann Brains. *Physics Letters B* 749: 514-518

[14] Norton, J. (1994). The Theory of Random Propositions. *Erkenntnis* 41: 325-352

[15] Norton, J. (2010). Probability Disassembled. *British Journal for the Philosophy of Science* 58: 141-171

[16] Norton, J. (2010). Cosmic Confusions: Not Supporting Versus Supporting Not. *Philosophy of Science* 77(4): 501-523

[17] Norton, J. (forthcoming). A Demonstration of the Incompleteness of Calculi of Inductive Inference *British Journal for the Philosophy of Science*

[18] Page, D. (2006). Is our universe decaying at an astronomical rate? *Physics Letters B* 669: 197200

[19] Rinard, S. (2013). Against Radical Credal Imprecision. *Thought: a Journal of Philsophy*, 2(1): 157-165

[20] Steinhardt, P., and Turok, N. (2002). A Cyclic Model of the Universe. https://arxiv.org/abs/hep-th/0111030

[21] Van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Clarendon Press.

[22] Vilenkin, A. (2006). Freak observers and the measure of the multiverse. *Journal of High Energy Physics*. DOI: 10.1088/1126-6708/2007/01/092

[23] Weatherson B. (2003). Are You A Sim? *Philosophical Quaterly* 53(212): 425-431

[24] Weatherson, B. (2007). The Bayesian and the Dogmatist. *Proceedings of the Aristotelian Society* 107: 169-185

[25] Weinberg S. (1987). Anthropic Bounds on the Cosmological Constant. *Physical Review* 59(22): 2607-2610