Why I Am Not A Boltzmann Brain

Version of July 15th 2024

[acknowledgements added Sept 5]

Sinan Dogramaci & Miriam Schoenfield

Forthcoming in *Philosophical Review*

Abstract: We give a Bayesian argument showing that, even if your total empirical evidence confirms that you have zillions of duplicate Boltzmann Brains, that evidence does not confirm that you are a Boltzmann Brain. We also try to explain what goes wrong with several of the sources of the temptation for thinking that such evidence does have skeptical implications.

1. What Is the Boltzmann Brain Skeptical Challenge?

Physicists tell us that a very strange cosmological model of our universe fits well with our current scientific evidence. On this model, after the stars and planets all decay away, the universe will continue to exist for so long that freaky events will be highly likely to eventually happen. In particular, although it's at any moment highly unlikely for the random movements of atoms in space to suddenly take the form of a fully functioning human brain, this is likely to eventually happen, and in fact likely to happen zillions of times, so many times that there will be many such "Boltzmann Brains" that, during their momentary existences, have experiences just like your present experience.

Let's call a universe that is described by such a model a "Boltzmann Brain Universe"—or BBU. The characteristic feature of a BBU that we're interested in here is just that it contains so many Boltzmann Brains—BBs—that not only will the total number of BBs vastly outnumber the number of ordinary observers like ourselves—OOs—but there will even be vastly more BBs than OOs having your exact present experience.[1]

[1] In certain long-lived universes, the random movements of atoms might form not only a short-lived radically deceived disembodied brain, but a whole stable galaxy that contains people that live happy lives and have mostly reliable beliefs about their general environments. We use the term "Boltzmann Brain"/"BB" only for radically deceived brains, and we use the term "Ordinary Observer"/"OO" to refer to observers who are not massively deceived with regard to their nearby surroundings. In the cosmological models that pose the most interesting skeptical challenge, the vast majority of brains are deceived. Our goal is to vindicate the reliability of our perception from purported skeptical threats posed by such recent cosmological models. It's certainly not our goal here to vindicate any claims about the origins of our galaxy.

Our current scientific evidence doesn't give conclusive support to any one particular cosmological model. And although a BBU model fits our current evidence, our future evidence may well rule it out. But, as Sean Carroll reports, our "current best-fit model for cosmology" is one that "will arguably give rise to a large number of BBs".[2] This raises an epistemological question. Supposing that we do obtain evidence that, by all ordinary standards of scientific reasoning, strongly confirms that we live in a BBU, should we take that evidence at face value?

The apparent difficulty with taking at face value any evidence that appears to indicate that we live in a BBU is, of course, that it generates a skeptical challenge. Since you and many of the BBs have indistinguishable experiences, how could you know you're an ordinary person observing real hands and a real computer (or real ink and paper), rather than one of the many BBs that will have a hallucination that is exactly like your current experience?

The challenge is powerful. In certain ways, it appears to be stronger than classical arguments for skepticism, which can only appeal to the metaphysical *possibility* of indistinguishable deceptions. But physicists have developed respectable theories according to which there *actually* are (or will be) BBs with experiences perfectly indistinguishable from our own.

In certain other ways, however, the challenge is weaker than classical arguments for skepticism. The classical argument appeals to a possible case, not an actual case, but the case is one in which *you* are deceived, and skeptics and anti-skeptics can all agree such a possible case exists. But it would beg the question for any skeptic to claim that there are actual cases in which you are deceived. So, how does the argument for skepticism go if it makes any useful appeal to empirical evidence that our universe is a BBU?

One way the skeptical argument could go is like this. The skeptic can propose a two-step argument for their skeptical conclusion under the supposition that we do obtain empirical evidence that, by ordinary standards of scientific reasoning, strongly confirms a BBU model.

[2] See Carroll (2021, p.7). Carroll's paper gives an expert's summary of the current science, and cites further relevant scientific literature. (We borrow the "BB" and "OO" acronyms from Carroll.) A recent New York Times article gives a snapshot of how the relevant evidence is inconclusive and rapidly changing: https://www.nytimes.com/2024/04/04/science/space/astronomy-universe-dark-energy.html.

Step 1: Make the ordinary scientific inferences that would lead to acceptance of a model on which our universe is a BBU, a universe populated almost entirely by zillions of BBs, including many with experiences indistinguishable from yours.

Step 2: Use some indifference reasoning to conclude that, given this model, you're far more likely to be a BB than an OO.

The first step has you infer that the universe's brain population, even the sub-population consisting of brains with your total evidence, is almost all BBs. The second step has you treat yourself as a random sample from this population, and so you infer you are almost certainly a BB. This two-step argument is how we'll understand the skeptical challenge posed by any empirical evidence for BBs.

Any response to the challenge posed by BBs must reject either of these two steps. Carroll (2021) hangs on to anti-skepticism by refusing to take the first step: he won't accept a cosmological model that includes zillions of BBs, even if we've obtained evidence that would appear to strongly confirm such a model.[3] Dogramaci (2020) hangs on to anti-skepticism by rejecting the inference in the second step: he won't infer that his experiences are a BB's experiences from the premise that most experiences that are just like his are a BB's experiences.

Both Carroll and Dogramaci, at some crucial point in each of their replies, make an appeal to the idea that coming to believe you're a deceived BB rather than a reliable OO, via the above two-step argument, is "cognitively unstable". How can you rest the conclusion that you're a BB on evidence that, according to that very conclusion, was deceptive evidence? But a number of other papers have made compelling critiques of the instability idea.[4]

Regardless of the merits of those critiques, our aim in this paper is to offer a response to the skeptical challenge posed by BBs that is simpler and more straightforward than appealing to cognitive instability. On our proposed view, there is no instability in believing, on the basis of our evidence, that we live in a BBU. The form of our proposal will be this: rational scientific reasoning is, and is nothing more than, Bayesian updating, and a Bayesian update on our

[3] He does this by insisting that the prior probability for any such model must be low enough that, even after we obtain any confirming evidence, the posterior remains extremely close to zero. See Carroll (2021, p.17).
[4] Kotzen (2021) and Avni (2023) make some compelling critiques of the instability idea. Wallace (2023), though, gives a way of arguing for Carroll's desired view (that our evidence does not confirm any model that implies that our evidence is unreliable) in a way that may help it avoid some of these objections. We will discuss Wallace below.

scientific evidence—even supposing the evidence strongly supports the hypothesis that we live in a BBU—leaves anti-skepticism perfectly stable and perfectly intact. There is no skeptical threat coming from Boltzmann cosmology, and there is no threat to Boltzmann cosmology coming from skepticism.

2. A Bayesian Solution to the Challenge of Boltzmann Skepticism

A Bayesian approach to evaluating the challenge is especially natural since the challenge concerns the interpretation of empirical evidence and Bayesianism is a leading theory of scientific reasoning, ie. of how it's rational to respond to empirical evidence.

We're going to assume, for the purposes of this paper, that the traditional skeptical challenge has been addressed. That is, we're assuming that, at least until you learned anything about modern cosmology, you were justified in being highly confident that you are a reliable ordinary observer and not a brain produced moments ago through the random fluctuation of particles.

Let E be your *total* empirical evidence. The Bayesian approach models the acquisition of empirical evidence as the updating of a prior probability function, so we'll suppose that you have a so-called ur-prior, Pr, which is your rational probability function prior to E, your total empirical evidence. We assume that, independently of any empirical evidence that there are zillions of BBs, anti-skepticism is rational, and we'll model that assumption by supposing that the ur-prior, Pr, initially assigns a high probability to the claim that we are reliable ordinary observers, OOs.[5] Let's also just suppose—for the sake of confronting the intuitively strongest threat of skepticism—that E does strongly confirm the hypothesis that we live in a BBU, that is, E significantly raises the probability of a cosmological model in which the OOs are vastly outnumbered by, as we're putting it, "zillions" of BBs.[6] What we will see is that this concession makes no difference to the argument we are going to make that our empirical evidence does not support skepticism.

---

[5] Some anti-skeptics, like those who favor an IBE approach, require that you first have some suitably coherent experiences before you can be confident that you're an OO. Those anti-skeptics can still accept the arguments of this paper, taking Pr to be your credence function prior to learning whatever evidence indicates that you live in a BBU, and letting E be the empirical evidence that supports the BBU cosmological theory. (Even so-called dogmatists, who think that the epistemic significance of perceptual experience is what *makes* anti-skepticism justified, can agree that the prior probability that my experiences are or will be reliable must be high, ie. that Pr(OO) must be high. See Pryor (2013) for discussion.)

[6] Later in the paper we'll discuss how some views deny that E confirms that there are zillions of BBs.

What you're ultimately really worried about, when confronting the skeptical challenge posed by evidence that we live in a BBU, is not really the probability of any cosmological model, but rather the probability that *you* are a BB. Put in Bayesian terms, and in the first-personal language that each of us would use, the thing I'm really worried about is this: how likely it is that I'm a BB given E, that is, how high is the posterior probability, Pr(I'm a BB|E)? We'll just write that more briefly as Pr(BB|E). (Throughout, in probability statements, we'll often use "BB" instead of "I'm a BB", and "OO" instead of "I"m an OO", for brevity.)

In evaluating Pr(BB|E), we can assume that the hypothesis that I'm an OO is the only live alternative, consistent with E, to the hypothesis that I'm a BB. (Any alternative third hypothesis consistent with E would be some other skeptical hypothesis such as: I'm the victim of Descartes' evil genius. And recall that, in facing down this new skeptical challenge posed by the evidence for BBs, we're going to assume those traditional skeptical hypotheses have already been assigned negligible credence by some traditional anti-skeptical philosophical considerations.) So, we're interested in Pr(BB|E), and we're assuming that Pr(BB|E) and Pr(OO|E) sum to 1, that Pr(OO) is extremely high, and Pr(BB) is extremely low. And, now, Bayes' theorem relates all these together by the so-called Bayes factor, which is the ratio of the so-called likelihoods[7], Pr(E|BB) and Pr(E|OO), as follows:

$$\underbrace{\frac{Pr(BB|E)}{Pr(OO|E)}}_{\text{(posteriors)}} = \underbrace{\frac{Pr(BB)}{Pr(OO)}}_{\text{(priors)}} \times \underbrace{\frac{Pr(E|BB)}{Pr(E|OO)}}_{\text{(Bayes factor)}}$$

What we ultimately want to know is the leftmost ratio, the ratio of the posterior probabilities (which will also tell us the exact posterior probabilities, since they partition the live possibilities). We want to know whether it is a high ratio (skepticism), or a low ratio (anti-skepticism). We already know that the middle ratio, Pr(BB) / Pr(OO), is very low, because we are anti-skeptics, at least prior to the empirical evidence indicating that there are zillions of BBs. And what all this means is that there is just one way for the posteriors to end up in a high (skeptical) ratio: the Bayes factor, the ratio of the likelihoods, must be very, *very* high. It must be

[7] It's a confusing but entrenched custom to call the conditional probability of observable evidence, conditional on a theoretical hypothesis, a "likelihood".

high enough to overturn the low prior ratio and produce a high posterior ratio. For example, if the priors had a ratio of 1:100 (roughly a 1% chance I'm a BB), then in order to make the posterior probability that I'm a BB higher than 50%, the likelihoods would have to have a ratio of 100:1. Or, in order to make the posterior odds that I'm a BB higher than 1:10 (roughly a 9% or higher chance, which is still a worrying result!), the likelihoods would have to have a ratio of at least 10:1, ie. it would have to be that Pr(E|BB) is at least ten times higher than Pr(E|OO). These are just illustrations to help give a feel for what the skeptical challenge from BBs demands of the likelihoods. Realistically, the prior probability, Pr(BB), isn't even remotely as high as 1%, but we won't argue over what it is exactly. Our main argument will rest on the claim that there is no reason to think the Bayes factor is high at all. There is no reason to even think that Pr(E|BB) is strictly higher than Pr(E|OO).[8] We'll next give concrete support to this claim by examining it in the light of two popular conceptions of what our empirical evidence consists in.

a. Theory #1: Evidence As Fundamentally Phenomenal

Let's think about the likelihoods using two specific theories of what our evidence consists in. The first theory says that *empirical evidence is fundamentally phenomenal*. Getting evidence is a matter of having an experience that is like this or like that. Verbally describing the phenomenal character of an experience can be hard, but usually we can adequately describe the relevant aspects of experiences that we base our scientific theories on. And it's such a description that will feature in a Bayesian analysis, since probabilities and conditional probabilities operate only on propositions.[9] The posterior, Pr(BB|E), will be conditional on a proposition that describes the relevant experiences. So, the question each of us is ultimately interested in then becomes something like this: what is Pr(I'm a BB|I'm having an experience as of an apple falling from the tree, hitting the ground, slowly rotting, and …)? The ellipsis, "…", would then go on to describe the empirical evidence that ultimately leads us to believe in the law of gravity and all

[8] Carroll (2021, secs. 1.4.1. - 1.4.2) provides a useful critical discussion of some earlier authors who considered the epistemic significance of likelihoods. But those earlier arguments do not consider the likelihood that we think is most relevant to the skeptical challenge—namely, the conditional probability of our empirical evidence given that *I'm* a BB. Rather, those arguments rest on claims about the conditional probability of our evidence given that *there are many* BBs. Moreover, we think those arguments give incorrect values to the likelihoods that they do consider, because the authors use indifference principles to assign the values. We will argue against such indifference reasoning below, starting in section 3b. (Carroll explicitly observes (p.12) that the likelihoods featured in those arguments are the ones that an advocate of "indifference principles" would endorse.)

[9] See Schellenberg (2018, ch.7.1) for a characterization of this view of evidence—though she denies that it is the correct view of perceptual evidence. Quine (1951, sec. 6) and Lewis (1996, p.553) seem to view evidence this way.

the other laws of physics, and leads us to make scientific predictions including that there will be zillions of BBs in the total history of the universe. Filling in the ellipsis would be impossible to actually do in practice, but we can still easily understand what it would involve.

So using our Bayesian setup, in order to answer our ultimate target question (what is Pr(BB|E)?), we just need to answer this: is Pr(I'm having an experience as of an apple … |I'm a BB) very much higher than Pr(I'm having an experience as of an apple … |I'm an OO)?

Consider first, then, the likelihood Pr(E|OO) on the phenomenal theory of evidence. What is the probability that I have these experiences I'm having as of an apple growing, falling, and rotting, etc, given that I'm a reliable ordinary observer? Well, honestly, it's hard to say exactly. Whatever it is, it's something very low. An OO is in causal contact with an environment that they will reliably veridically represent (at least in their perceptual judgments, even if their phenomenal experiences don't have accuracy conditions). But, an OO still might undergo any one of zillions of possible experiences. It thus seems like Pr(E|OO) is one in however many zillions of possible experiences an OO might have. Perhaps some anti-skeptics would want to argue that our actual experience is something fairly normal and, therefore, our actual experience is relatively likelier than other stranger experiences that an OO could possibly have, and, therefore, normal experiences like ours have a higher than average probability for an OO. But we aren't inclined to argue in this way. We will be more generous to the skeptic and assume Pr(E|OO) is the average of the probabilities of all the zillions of possible experiences for an OO, ie. Pr(E|OO) indeed is one in however many zillions of possible experiences an OO might have. So, Pr(E|OO) is some extremely small number, and it certainly won't secure an anti-skeptical posterior just on its own.[10]

But, while Pr(E|OO), the denominator in our Bayes factor, considered just on its own hasn't offered us much anti-skeptical comfort, let's now also consider the numerator, Pr(E|BB). It's again hard to say exactly what it is. It's clearly something very low, though it's not clear exactly how low. But—the important thing now—we can argue that it's *lower* than Pr(E|OO). We

[10] One might try to argue as follows: Pr(E|OO) is *extremely* low, lower than the average of the probabilities of the zillions of possible experiences an OO might have, and even lower than the vanishingly small Pr(E|BB). Why? Because, one might argue, supposing I'm an OO, it's very unlikely that I'd get evidence E—for that would constitute misleading evidence in support of the claim that in fact I'm a BB.

Of course, this line of argument requires thinking that Pr(BB|E) is high—a claim that we're in the midst of arguing against. But later, in section 3, we will critically examine a view, one called Center Indifference, that appears to support this assumption. We'll argue both that the view is false, and that upon closer examination the view—contrary to what even some of its advocates believed—turns out to be inconsistent with Pr(BB|E) being high.

can argue that it must be lower because a BB, the product of random processes, undergoes a random sequence of experiences. The range of possible experiences a BB could have is much greater than the range possible for an OO. While an OO's experiences must be reasonably coherent since they must (produce perceptual judgments that) reliably veridically represent a real possible environment that the OO could exist in, a BB's experiences are totally unconstrained. A BB could experience anything an OO could, and more. While OOs like ourselves might occasionally undergo illusory experiences as of impossible scenes (the famous waterfall illusion is one standard example), in order to count as being an OO, you must be reliably hooked up to the world around you, and that puts constraints on what your experiences could be like. (If all you ever experienced were waterfall illusions, you wouldn't be an OO.) A BB, in contrast, could have one unimaginably nightmarish jumble of incoherent sensations after another. It follows that there are more experiences a BB could have than there are experiences an OO could have. Furthermore, we have no reason to suppose any one possible experience (coherent or incoherent) is any more likely than any other for a BB to undergo. BBs exist only because atoms will, in the long run, randomly assemble themselves in any possible configuration. And since we have no reason to suppose any experiences are more likely than any others to be tokened in the possible configurations of a functioning brain, we can suppose each of the bazillions of possible experiences that a brain could have has a one-in-a-bazillion chance of occurring in a BB. So, Pr(E|BB) is extraordinarily low, and—again, the important thing—lower than Pr(E|OO). This means the Bayes factor is less than 1, and the prior ratio, Pr(BB) / Pr(OO), will only go down when it is updated into the posterior ratio, Pr(BB|E) / Pr(OO|E). Thus, the evidence E does not increase the probability that I'm a BB, and so does not support skepticism.

In the preceding argument, we made some claims of the form "we have no reason to suppose…". This is not a weakness that undermines our argument. It's not a weakness because we are assuming that anti-skepticism is the rational default attitude to start out with in our priors, and there must be some clear reason to overturn that attitude if there is any threat of skepticism. If we have not been given, and cannot find, any reason to think there is a threat, then there *is* no threat.

To make the preceding argument, we focused on the coherence of our evidence. We said that there's no reason to think I'm more likely to get such coherent evidence if I'm a BB than if I'm an OO. Should we now be worried about another hypothesis, the hypothesis that I'm a BB

whose experiences are coherent? The worry can seem compelling because even if we restrict attention to that small proportion of BBs whose experiences are coherent, there are still more such BBs than there are OOs (supposing that I live in a BBU). And what about the hypothesis that I'm a BB having exactly this total sequence of experiences? That hypothesis might also seem worrisome in light of the thought that (in a BBU) there are more BBs than OOs having exactly this total experience.

Our reply is that these specific skeptical hypotheses may well be confirmed, but that is no confirmation of the hypothesis that I am deceived, and so there is no support here for skepticism. Our main argument already showed that the skeptical hypothesis that I'm a BB is not confirmed. This is consistent with the possibility that various highly specific skeptical hypotheses undergo a boost as our evidence transforms our priors into our posteriors. (Even though any specific skeptical hypothesis logically entails the general skeptical hypothesis that I'm deceived, confirmation of the former is not confirmation of the latter. See section 3a below for more discussion of this kind of intransitivity exhibited by the probabilistic confirmation relation.) This kind of thing happens every time we have an ordinary experience on the phenomenal conception of evidence. When I have an experience as of a cup of coffee, the specific skeptical hypothesis that I'm deceived *and* hallucinating coffee is confirmed.[11] On the phenomenal conception of evidence, my experiences always confirm the super-specific hypothesis that I'm deceived yet having all the exact experiences I'm actually having. Nevertheless, what we've argued is that the general skeptical hypothesis that I'm a BB is not confirmed, and the anti-skeptical hypothesis that I'm an OO is not diminished. And this remains true even while we've granted to the skeptic that our evidence may also strongly confirm that I live in a BBU populated by zillions of BBs—this is because that concession doesn't affect the likelihoods that ensure that E doesn't confirm that I'm a BB (a point that we'll keep returning to). This suffices to answer the skeptical threat that I'm a BB posed by my empirical evidence, at least on the phenomenal conception of evidence.

[11] See White (2006) for one very clear discussion of this. Relying on the phenomenal conception of evidence, he correctly observes that such specific hypotheses must be confirmed, ie. their probability must be raised.

b. Theory #2: Evidence as the Propositional Content of Experience

The other theory of evidence we'll consider is the view that our perceptual experiences have propositional contents, and these contents are our empirical evidence.[12]

When, say, you see an apple, what is E in that case, ie. what is the propositional content of your experience? And what is your duplicate BB's evidence?

On some externalist theories of evidence, an OO seeing an apple gets different propositional evidence from their internal duplicate BB who is hallucinating an apple.[13] But the skeptical challenge posed by BBs isn't forceful for such externalist theories. The skeptical worry that you might be a BB rests on the assumption that your evidence is no different and no better than a BB's evidence—an internalist assumption. We accept the internalist assumption that any OO has exactly the same total evidence as their duplicate BBs each have, so we will respond to the skeptical threat on these terms, terms that favor the skeptic's side.[14]

Suppose, then, that when you see an apple, your evidence is the proposition that we'd express with the sentence "there is an apple", and your duplicate BB's evidence is the same proposition. Now consider the resulting posteriors, priors, and Bayes factor:

$$\frac{Pr(BB|there\ is\ an\ apple)}{Pr(OO|there\ is\ an\ apple)} = \frac{Pr(BB)}{Pr(OO)} \times \frac{Pr(there\ is\ an\ apple|BB)}{Pr(there\ is\ an\ apple|OO)}$$

Is the Bayes factor larger than 1, and is it larger by enough so as to turn anti-skeptical priors into skeptical posteriors?

If "there is an apple" is just a bare existential generalization that implies nothing about where or when the apple is located, then our question becomes, how likely is it for an apple to exist in the universe (somewhere, sometime) if I'm a BB, and how much more or less likely is it for an apple to exist if I'm an OO who reliably veridically perceives whatever is in my local environment? Worryingly, the supposition that I'm a BB arguably raises the probability that the universe produces lots of other random stuff too, and so the likelihood Pr(there is an apple somewhere, sometime|BB) might be relatively high. And, to continue the worrying line of thought, it's not clear that Pr(there is an apple somewhere, sometime|OO) is particularly high,

[12] See Pryor (2000), Miller (2016), and Comesana (2020) for some defenses of this view.
[13] See Williamson (2000, ch.9) for a classic example.
[14] See Saad (forthcoming) for one kind of radically externalist response to the skeptical challenge from BBs.

much less high enough to offset the skeptical damage done by the numerator. The exact numbers here are hard to specify, but it would be much more comforting to us anti-skeptics if we could somehow reject this whole worrisome line of thought.

Fortunately, we can reject it. We can reject it because, on entirely independent grounds, it is implausible that the empirical evidence we get from perception is such a bare existential with no implications about time or location. Two arguments show that our evidence has content that concerns the perceiver's own present time and location, an *egocentric* content as we'll call it.

The first argument is that our empirical evidence serves as a basis for our rational actions, but it can make our actions rational only if it gives us information about our present local circumstances. If it can be rational to reach your hand forward on the basis of your perceptual experience of an apple, then the experience must put you in a position to rationally think there is an apple in your reach, but the bare existential that there is an apple somewhere can't help with that task at all. Your experience must somehow give you evidence not just that *there is* an apple, but that an apple *is there*. What you experience is that the apple is there in front of you, an egocentric content.[15]

The second argument is that empirical evidence can serve as the basis of standard rational inductive inferences only if that evidence concerns, or implies things about, your *own* observations. (We thank [OMITTED] for showing us this.) Suppose you know an urn contains exactly three marbles, each of which is red or blue. You know at least one of them is red, and at least one of them is blue. Suppose also that you know that other people have taken very many samples from the urn (draws with replacement). You don't learn the details about what they observed, but you're sure that at least one of their observations was of a red marble and at least one of their observations was of a blue marble. Now you shake the urn, draw a marble from it for yourself, you observe that it's red, and you replace it. Your own observation confirms that most of the marbles in the urn are red. You've empirically confirmed this. (If, for example, your prior probability that most of the marbles are red was ½, then, by Bayes' theorem, your posterior should be boosted to ⅔) But the content of your empirical evidence, ie. the content of your observation, cannot be the proposition *there is a red marble*. It cannot even be *a red marble was observed*. These claims don't boost your confidence that both marbles are red. That's because

[15] See Perry (1979) for a classic presentation of the basic idea, though Perry is focused on the contents of beliefs rather than experiences. See Siegel (2021, sec. 3.4) for a recent survey of the considerations supporting the idea that perceptual experiences have egocentric content (or 'indexical content' as she calls it).

you already knew that there is a red marble and you already knew that a red marble was observed (and so, in Bayes' theorem, the posteriors do not change from the priors). This means that the content of your observations must be something stronger than bare existential claims about a red marble or even about an observation of a red marble. (Suppose also that you're drawing from the urn in a blank room with no clock, so your evidence cannot be that a marble was sampled at, say, noon and it turned out to be red, or anything like that.) What this case is designed to show is that, when you draw a red marble, and when you gain empirical evidence in general, your observation must have an egocentric content.[16] When you observe the red marble that you drew, you must be empirically learning that there is a red marble *here* and *now*, in *this* observation *of yours*. That is the kind of antecedently unknown evidence that has a lower likelihood if the urn has a non-red marble and a higher likelihood if the urn has two red marbles, and thus that is the kind of evidence that can confirm the two-red hypothesis, and so, we conclude, that must be the kind of evidence you're obtaining.

So, empirical evidence has egocentric content. And if such content is what our empirical evidence consists in, then we can address the skeptical challenge from BBs. Now the Bayes factor looks something like the following, with our total egocentric empirical evidence going in on the left sides:

$$\frac{Pr(here\ is\ an\ apple, this\ apple\ is\ sweet, here\ is\ a\ hand, now\ it's\ raining, \ldots | BB)}{Pr(here\ is\ an\ apple, this\ apple\ is\ sweet, here\ is\ a\ hand, now\ it's\ raining, \ldots | OO)}$$

As usual, exact numbers are hard to give, but now we have clear and strong reasons to think the numerator will be much lower than the denominator. The denominator itself is, admittedly, some very low number—the circumstances we actually perceive ourselves to be in are ordinary enough circumstances, but there are very many possible ordinary circumstances an OO could be in, and what were the chances an OO would land in these *particular* ordinary circumstances? However, large as the range of ordinary possible circumstances of an OO is, there are additional possibilities that a BB could be in. And there is one particular possibility that a BB is

[16] We think it's most natural to include concepts of oneself in the content: *this* observation *made by me, here, and now,* is of a red marble). Others may prefer to say it is a pared down but still *de re* content: *this* observation is of a red marble. That view would work for our purposes too. All we insist on ruling out is the view that the content is purely *de dicto*: *an* observation was/is of a red marble.

overwhelmingly likely to be in: while a BB *could* randomly form in space *along with* an apple, hands, rain, and everything else that our perceptual experiences represent, a BB is most likely to exist with nothing around it to be perceived at all, nothing but the void of empty space around it. As unlikely as it is for a BB to form at all, it's drastically more unlikely for additional things to simultaneously form around it.[17] And even if additional things did form around the BB, it is not especially likely they will be the kind of stable and sensible objects a brain could even perceive. Therefore, the evidence we have is more likely supposing we are OOs than supposing we are BBs.

We've now completed our argument for the claim that there is no reason to believe we are BBs or to doubt we are OOs.

Notice again that, throughout the whole argument, we left in place an initial concession: we imagined that my total evidence strongly confirms a cosmological model with zillions of BBs. Our analysis shows that, even conceding that my evidence strongly confirms that I live in a BBU, it does not at all confirm that I'm a BB. And the reason why this is so is, again, that the likelihoods relating our evidence, E, to these hypotheses are very different. It can be that the likelihood Pr(E|I live in a BBU) is as high as you want, and that it is is higher than the likelihoods for all the competing cosmological models—that concession doesn't interfere at all with the fact of the very low probability for the likelihood Pr(E|BB), ie. for Pr(E| *I am* a BB). It is only the likelihoods of Pr(E|BB), together with Pr(E|OO), that bear on the question of skepticism.

What does this analysis imply about the two-step argument for Boltzmann skepticism that we described early on? The argument was this.

> *Step 1:* Make the ordinary scientific inferences that would lead to acceptance of a model on which our universe is a BBU, a universe populated almost entirely by zillions of BBs, including many with experiences indistinguishable from yours (and thus, we've even conceded, many BBs who have the same total evidence as you have).
>
> *Step 2:* Use some indifference reasoning to conclude that, given this model, you're far more likely to be a BB than an OO.

[17] See Carroll (2021, pp.9-10).

We said that Carroll (2021) claims the first step must be irrational, and Dogramaci (2020) claims the second is. In light of our Bayesian analysis, we take sides with Dogramaci and not Carroll. We see no problem with the inferences in step 1; if the physicists were to tell us that the evidence appears to, by ordinary scientific standards, strongly confirm a BBU cosmological model, then we'd happily and rationally infer that we live in a BBU. It's the inference in step 2 that would be irrational. Dogramaci gave his own explanation of why the second inference is irrational (explaining it partly by applying Carroll's notion of cognitive instability in a different way than Carroll does). We hope that we've here given an additional satisfying explanation of why the second inference is irrational, this time explaining it in purely Bayesian terms. Let E again stand for our total empirical evidence. The second inference is mistaken because, as we've shown, even if Pr(I live in a BBU|E) is high, nevertheless Pr(I'm a BB|E) is low, and the explanation why it is low is given by Bayes' theorem together with the various explanations we've given of the values that go in the Bayes factor.

3. Why Evidence of Boltzmann Brains Tempts Us into Skepticism

In the rest of the paper, we turn to diagnosing why we are tempted by the mistaken thought that skepticism follows from scientific evidence indicating that there will be zillions of BBs. We'll offer two (mutually compatible) explanations of why we make this mistake.

a. First Explanation: The Allure of Probabilistic Fallacies

Why does, or did, Boltzmann skepticism tempt us? Here is one explanation.

First, let's suppose that

1. E confirms that I live in a BBU,

or in probabilistic terms:

1. Pr(I live in a BBU|E) > Pr(I live in a BBU).

Let's also suppose, and in fact we will even argue, that

2. That I live in a BBU confirms that I'm a BB,

or, again restating things, in probabilistic terms:

2. Pr(I'm a BB|I live in a BBU) > Pr(I'm a BB).

(A simple argument for (2) is that

3. Pr(I live in a BBU|I'm a BB) > Pr(I live in a BBU|~I'm a BB).

We don't think the left-hand side of this inequality is a high probability—a claim we'll actually return to again below—but it's surely higher than the right hand side. So, we think (3) is an intuitively true inequality, and (2) follows from (3) by Bayes' theorem.[18])

We propose that (1) and (2) appear to pose a skeptical threat because from them we are tempted to infer, though wrongly, that the scientific evidence E confirms that we're BBs. We're tempted to infer that skeptical conclusion because we are tempted to treat the confirmation relation as transitive. If the confirmation relation *were* transitive, (1) and (2) *would* constitute a good basis for an argument that the scientific evidence confirms that we're BBs. And since there is also a common tendency to confuse a hypothesis being confirmed (ie. the posterior is high*er* than the prior) with the hypothesis being made likely (ie. the posterior is high)[19], we also have here a possible explanation of why people are tempted to wrongly think the scientific evidence threatens to make it likely that we're BBs.

But the scientific evidence does not even confirm, let alone make likely, that we're BBs. As Bayesian analysis famously reveals, the confirmation relation is intransitive, even though there is a deep and persisting human intuition that it is transitive.[20] (It just *sounds* intuitive, even

[18] An additional argument for 2 could be given if one accepted a principle that we'll define in the next section, Center Indifference. But we reject that principle. Our only argument for 2 is 3. (2 follows from 3 by properties (i) and (ii) of Bayesian confirmation that are given in Appendix B.)

[19] Like when a positive test for a rare disease scares you. See Crupi, Fitelson, and Tentori (2008), and references therein, for some interesting discussion of the empirical psychological hypothesis that ordinary assessments of probability are often actually guided by assessments of confirmation relations.

[20] See Kotzen (2012) for a valuable discussion. A simple illustration of the phenomenon: that the roulette wheel landed 1 or 2 confirms that it landed 2 or 3, and that it landed 2 or 3 confirms that it landed 3 or 4, but obviously that the roulette wheel landed 1 or 2 does not confirm (in fact it rules out) that it landed 3 or 4. If we want a simple

when put in schematic terms: if F makes it likelier that G, but E makes it likelier that F, then surely E makes it likelier that G! Formulating it with counterfactual conditionals about rational learning makes it even more irresistible: if learning F would make a rational person in my position more confident that G, and learning E would make a rational person in my position more confident that F, then surely learning E should make a rational person in my position more confident that G!)

The tendency to treat the confirmation relation as transitive (when it is not) is related to the tendency to neglect part of your total evidence. If your total evidence *were* just that there are zillions of BBs, then as (2) correctly says, this *would* confirm that you're a BB (in the sense of raising the probability). But your total evidence includes all of E, and, as Dogramaci (2020) says and as we agree and hope we've helped to show here, that total evidence does *not* confirm that you're a BB, in the sense that it does not raise the probability at all.

b. Second Explanation: The Allure of Indifference Reasoning

The second step in the two-step argument for Boltzmann skepticism moves from the premise that I live in a BBU to the conclusion that I'm almost certainly a BB. This step is justified by a principle of indifference over so-called "centered worlds". A centered world is a time and place associated with a possible world. Subjects occupy centered worlds, have credences about which centered world they occupy, and can gain evidence that bears on which one they occupy. A principle of indifference over centered worlds was first proposed by Elga (2004). Weatherson (2005) and Builes (forthcoming) each add some clarifications. Weatherson opposes the principle; Builes endorses and defends it, and he applies it to the threat of Boltzmann skepticism. Builes, calling centered worlds in the same possible world "similar", formulates the principle like this:

> *Center Indifference* (**CI**): for any two similar centered worlds $c1$ and $c2$, if both $c1$ and $c2$ are compatible with your evidence, then it is rationally required to set $Cr(c1|c1 \text{ or } c2) = ½$. (Builes, forthcoming, p.3)

example where none of the hypotheses entail or refute any of the others (a feature of the hypotheses in our main discussion), consider the hypotheses that the wheel landed in these intervals: 1-15, 5-20, and 10-25.

This implies that, in general, if your evidence is consistent[21] with a number of centered worlds that are all in the same possible world, then you should give equal credence to your occupying any of those centered worlds.

The intuitiveness of the threat of Boltzmann skepticism depends on the intuitiveness either of **CI** or at least of the general idea of indifference that lies behind a precisification like **CI**. (Builes and Carroll both explicitly recognize the role of **CI** or some such indifference principle in the intuitive threat of Boltzmann skepticism.[22]) **CI** helps make the threat of Boltzmann skepticism intuitive because **CI** validates the second step in the two-step argument for Boltzmann skepticism. For, **CI** immediately implies that Pr(I'm a BB|E & I live in a BBU) is extremely high, for any E consistent with my being any one of many BBs.

CI has this implication on both theories of evidence that we discussed.[23] However, for the duration of the critical but charitable examination of **CI** that we want to have now, we'll assume the phenomenal theory of evidence. This is for the following reason. The main motivation for **CI** is that you should be indifferent between "indistinguishable" centers, including those of your duplicate brain-in-a-vat or your duplicate BB-in-the-void.[24] But if my evidence consisted of

---

[21] Without giving a comprehensive theory of the logic of egocentric contents, we'll assume (for the sake of avoiding the trivialization of **CI**) that two egocentric contents can be *inconsistent* in an intuitive sense. When I view a red ball, my experience's egocentric content is consistent with the content of the experience of other centers that also view a (similar looking) red ball, but *inconsistent* with the content of the experience of centers that view, say, a blue ball or a red cube. The fact that my experience is *about me* and the blue ball viewer's experience is *about them* does not, we assume, make the two contents consistent.

[22] Carroll (2021, p.12) says that the assumption of some "indifference" principle is a part of standard analyses of cosmological models that say we live in a BBU. His rough pass at the the content of the principle is, "given some reference class of intelligent observers, we are equally likely to have been any of them, and should reason accordingly"; see p.12, citing earlier authors that he attributes this to.

Builes (forthcoming, sec. 7) takes **CI** to raise a serious skeptical threat when it is combined with the premise that most minds like yours are Boltzmann Brains. He suggests that one attractive way to avoid the skeptical threat is to endorse presentism and a version of **CI** that only requires indifference over centered worlds that (presently) exist.

[23] In a BBU, most centers in which things phenomenally seem as if there's a computer sitting on a desk in front of me are BBs, and in fact it's also true that, in a BBU, most centers in which there is a real computer and desk and whatnot in front of me are also BBs—BBs who are deceived about what's happening beyond their immediately perceived present circumstances.

[24] See Elga (2004, p.383) and Builes (forthcoming, p.2) for explicit descriptions of the main motivating thought experiments as ones that involve "indistinguishable" centers. Builes concedes an objection that Weatherson (2005, sec. 3) makes against Elga: if a subject in one center can possess evidence that rules out their being in another "indistinguishable" center, such a subject need not be indifferent across the two centers. Builes endorses the formulation of **CI** we stated above, a formulation which does not demand indifference in such a case, and he argues that **CI**, so formulated, still enjoys other motivations (that concern intuitions about disagreement between evidence-sharing peers; see p.4). But as Builes recognizes (pp.12-3), **CI** supports our skeptical intuitions about the possibility of being a BIV or BB-in-the-void only if your evidence is consistent with such possibilities. Builes thus adopts this assumption himself for the duration of his discussion of the relationship between **CI** and skepticism. As he says: "If you're an externalist about evidence, and you take the fact that you have hands to be part of your

external world propositions that are the content of my experience, and we're working in a classical Bayesian framework, it would follow that I should be certain of those external world propositions regardless of what I know about other centers having indistinguishable experiences. Therefore, to interpret **CI** in a way that respects its motivations**,** we'll assume the phenomenal theory of evidence in this section.[25]

So, **CI** has some intuitive motivations, and **CI** validates the second step in the two-step argument, and this means the whole two-step argument enjoys strong intuitive support since the first step in the two-step argument is highly intuitive on its own. That first step just has us infer that we live in a BBU, at least in the event that we get empirical evidence that, at face value, strongly confirms that we live in a BBU. Given the intuitiveness of **CI**, then, both steps of the two-step argument for skepticism are intuitive.

This is a nice explanation of why the argument for Boltzmann skepticism is so tempting. But our goal is to explain why we are *wrongly tempted* by Boltzmann skepticism. What then explains why the skeptical argument is intuitive *but wrong*? Which step in particular, in the two-step argument, is intuitive but wrong?

As we've said, our view is that it's the second step that's wrong—the inference from "we live in a BBU" to "I'm a BB"—and we've argued for this by appealing to the likelihoods. What we will argue now is that, although **CI** appears to have skeptical consequences, and although **CI** also appears to be intuitively true, neither appearance is correct. First, we'll argue that in fact, **CI** doesn't have skeptical consequences, because **CI** turns out to entail (with a plausible auxiliary premise) just the result that it requires in order to block the *first* step of the two-step skeptical argument. Second, we'll argue that **CI** has counterintuitive commitments that show that it should ultimately be rejected.

We start by examining the following remarkable fact about **CI**: it implies that our evidence does *not* support the BBU hypothesis. We're going to give our own arguments for several versions of this result, and we'll examine their significance and explain how they expose deep problems with **CI**. But first let's consider two other arguments for the view that **CI** invalidates the first step, arguments that Carroll and Wallace have offered.

---

evidence, then none of these skeptical cases will be compelling. For the moment, then, let us assume an internalist conception of evidence." (p.13)

[25] There may be clever and complex ways **CI**'s motivations can be reconciled with the view that our evidence consists of propositions about the external world. We are just recommending the phenomenal view as the most charitable lens through which to think critically about **CI**.

As we mentioned in section 1, Carroll (2021, pp.17-18) is someone who refuses to take the first step. He refuses to infer from his evidence that he lives in a BBU.[26] He refuses because, if he did infer it, then by **CI** (or some principle like it) he would think he is surely a Boltzmann Brain, and that would imply that his original evidence is misleading. For the adherent of **CI**, inferring from empirical evidence that you live in a BBU is, in this sense, "cognitively unstable". Carroll concludes that the prior probability of any BBU cosmological model must be low enough that its posterior remains very low conditional on our scientific evidence.

We're going to raise a problem with this idea that "cognitive instability" explains why our evidence cannot strongly confirm a BBU cosmology. As we'll show below (in **Result 2**), a **CI** adherent like Carroll must limit the extent to which my empirical evidence confirms that I'm in a Boltzmann Brain universe *even under the supposition that I'm an ordinary observer*. (That is, for any body of evidence, E, Pr(I live in a BBU|E *& I'm an OO*) can only be so high.) But under the supposition that I'm an ordinary observer, there's nothing cognitively unstable about thinking that I live in such a universe. If cognitive instability were the explanation for why no possible body of evidence I could get would rationalize a high credence that I'm in a BBU, then under a supposition that removes the cognitive instability I *should* be able to have evidence confirming that I'm in a BBU. But as we'll show, the **CI** adherent can't say this. Cognitive instability, then, is not the culprit.

Wallace (2023) offers a different way to argue for Carroll's desired conclusion, the conclusion that Pr(I live in a BBU|E) must be low. Wallace shows how to prove this conclusion from a few assumptions and the laws of probability.[27] His argument relies on a distinction between what he calls "proximal" and "primary" evidence. "Proximal" evidence consists in the reportings (or my recollections of the reportings) of journals and textbooks as to what other scientists have observed, and "primary" evidence consists of propositions about the scientists' experimental outcomes themselves. Wallace's proof rests on two assumptions. Let E* be the proposition that the proximal evidence E, is "reliable as to what the primary evidence is" (Wallace, p.297). The first assumption is that Pr(E*|E and I live in a BBU) is very low—this formalizes the idea (which the **CI** advocate endorses) that the BBU hypothesis is

[26] Dyson, Kleban, and Susskind (2002, see esp. sec. 6 and the last sentence) are another notable group of physicists who have at least sympathized with the idea that, for the sake of avoiding the conclusion that we live in a BBU, we must revise our initial assessments of how likely various cosmological models are made by our evidence.
[27] He shows this for any theory, but we'll write in the theory we're interested in, the theory that I live in a BBU. Wallace says that he is neutral on whether his assumptions hold in any particular case.

self-undermining. The second assumption is an anti-skeptical assumption: Wallace assumes that Pr(E*|E) is very high, ie. the probability that, given our proximal evidence, we are representing the world reliably, is very high.

Wallace's proof is insightful (and inspired some of our own thinking here), but it has two limitations that are especially relevant to Boltzmann skepticism.

First, its soundness won't be accepted by the skeptical position targeted by this paper, a position that says we must become skeptics upon receiving E. Such a person will deny the proof's second assumption. Of course, every anti-skeptical argument has some premise that a committed skeptic will deny, and there is no simple recipe for deciding who is begging a question, and who bears a burden of proof. But in this case, we're considering a skeptical threat *posed by E*, and so it's notable that Wallace's proof shows otherwise (E does not confirm a skeptical hypothesis) only given the assumption that we are representing the world *reliably* conditional on E.

Second, although the factorization of a scientist's overall evidence into two parts, proximal and primary, makes sense for a wide range of realistic scientific cases, it does leave other cases unaddressed, for example the case where we just look and see a universe brimming with Boltzmann Brains. The case, though weird, seems possible, and it's unclear how it would involve three different propositions playing the requisite roles for Wallace's proof. If, for example, E is that it appears to me that the universe is brimming with BBs, and the theory in question is that the universe is brimming with BBs, then what is E*?[28] If the advocate of **CI** wants to show that your evidence in these examples does not really confirm the theory that you live in a BBU, it would be useful to them to have a proof that does not require an E* proposition.[29]

We'll now present several results that, without relying on the distinction between proximal and primary evidence, and without assuming that E is reliable empirical evidence, show that **CI** imposes significant limits on how much E can support the BBU hypothesis. The results will show that **CI** has some very odd consequences, which will lead us to conclude that **CI** must be rejected. The results will also illuminate the core problem with **CI** and help us

[28] E* certainly can't be the material conditional E ⊃ I-live-in-a-BBU, since that conditional is not unlikely given E and I-live-in-a-BBU, and it's at least unclear how a non-truth-functional conditional could express a proposition that does better at playing the E* role.

[29] The scope of Wallace's argument is acknowledged, at least implicitly, in the last two words of his title: "A Bayesian Analysis of Self-Undermining Arguments in Physics.

explain (in the next subsection) why we shouldn't accept the inference it recommends in the two-step argument.

To state these results, we'll use H_E for the hypothesis that the vast majority of brains with total evidence E are BBs. Then $\sim H_E$ is the hypothesis that some correspondingly smaller proportion (anything less than a vast majority) of brains with E are BBs. And let's again abbreviate the hypothesis that *I'm* a BB (or an OO) to just "BB" (or "OO") in our probability statements. Then **CI** immediately implies this:

Lemma: $Pr(BB|E\&\sim H_E) < Pr(BB|E\&H_E)$

(If, eg., we imagine H_E says at least 99% of brains with E are BBs, then **CI** says $Pr(BB|E\&H_E)$ is somewhere between 99 and 100%, and $Pr(BB|E\&\sim H_E)$ is somewhere between 0 and 99%.)

Our first result rests on just one more assumption in addition to **CI**, which is this:

Defeat: $Pr(H_E|BB) = Pr(H_E|BB\&E) < ½$

We think this is a very plausible premise, no matter what E is. In the presence of the defeating claim that I'm a BB, it is completely random what phenomenal experience I have; it's just a roll of the dice. So, E, a description of those randomly generated experiences, cannot confirm (or disconfirm) any cosmological hypotheses. The posterior probability of hypothesis H_E will then be, at least approximately, the same as its prior probability, some low value. In other words, $Pr(H_E) \approx Pr(H_E|BB) = Pr(H_E|BB\&E) < ½$. (We say "approximately" because the supposition that I'm a BB can itself boost H_E a tiny bit. After all, that I'm a BB is one data point in favor of a population of BBs.[30] Still, one data point does not make an antecedently wildly improbable hypothesis more than 50% probable. The fact that I'm a BB does not make it over 50% probable that 99.99999999% of brains like me are BBs.)

From **Lemma** and **Defeat**, and a few laws of probability, we can show this (the proof is in appendix A):

[30] This is why we said, in the previous subsection, that we think $Pr(H|BB) > Pr(H|\sim BB)$.

Result 1: $Pr(H_E|E) < ½$

To us, **Result 1** looks like a weird kind of scientific revisionism: from philosophical premises, we've proved that our scientific evidence cannot strongly confirm a cosmological model according to which most brains like our own are BBs. But to the adherent of **CI**, **Result 1** might seem to be just what they were looking for. **Result 1** blocks the first step of the two-step argument for Boltzmann skepticism. This was the step they were going to be forced to deny if they wanted to avoid skepticism (since **CI** validates step 2), but now the **CI** adherent has a way to argue that step 1 is mistaken. And the result does not assume any factorization of evidence into two parts, and the result also does not make the assumption (which the Boltzmann skeptic would consider question-begging) that E is reliable empirical evidence.

While it might be debatable what the significance of **Result 1** is, and even whether it's bad news or good news we will now give a second result that is unambiguously bad news for the **CI** adherent.

From the same assumptions, **Lemma** and **Defeat**, and a few laws of probability, we can show this (the proof is in appendix B):

Result 2: $Pr(H_E|E\&OO) < ½$

This is an absurd result. Our view is that it is so absurd that it overturns whatever intuitive support can be claimed for **CI**. Our results don't assume anything about what my evidence E is, so they hold for *any* body of evidence E and the corresponding hypothesis H_E. That means that it is impossible for any empirical evidence to support a cosmological model in which there are too many Boltzmann Brains that have that evidence *even on the supposition that I'm an ordinary observer*. If that is impossible, something must explain why. Carroll rightly appreciated that we need to find some explanation for such a strange phenomenon, so he appealed to cognitive instability. But the remarkable thing that **Result 2** shows is that I cannot gain evidence that makes probable a hypothesis about how many brains (with my evidence) are BBs *even supposing I am not one of the BBs*. There is nothing unstable about believing I live in a BBU under the *supposition* that I'm an OO. If I'm an OO, my evidence is trustworthy, but then

why can't I just look and see what the world is like, and—again, assuming I'm an OO—believe what my eyes tell me?

Can the adherent of **CI** escape **Result 2** by resisting **Defeat**, the only assumption it rests on other than **CI** (which implied **Lemma**)? We think not.

One reason is that **Defeat** is intrinsically very plausible, as we explained above. A more important reason is that, even if we drop the assumption **Defeat**, troubling results still follow just from **CI** alone.

Assuming only **Lemma** (which **CI** entails) we can prove this:

Result 2.1: $Pr(H_E|E\&OO) < Pr(H_E|E\&BB)$

(The proof is just the first 5 steps of the proof of **Result 2** in appendix B.[31]) This is another weird result. Again, it holds for any choice of E.[32] The reason **Result 2.1** is so weird is that, even if **Defeat** is false, the following equality is still highly plausible: $Pr(H_E|E\&BB) = Pr(H_E|BB)$. We take it that the adherent of **CI** cannot deny that; they cannot deny that the condition that I'm a BB neutralizes the force of my empirical evidence. But then **Result 2.1** implies this:

Result 2.2: $Pr(H_E|E\&OO) < Pr(H_E|BB)$

This result, again, holds for any body of evidence E. So, in particular, it holds when E is the proposition that it appears as if I've just observed some very large sample of brains with evidence E, and they are all BBs. Then the left side of **Result 2.2** concerns the probability of the hypothesis that most brains with E are BBs, on the supposition that I'm a reliable OO and I seem to have just observed a large sample that, by ordinary scientific standards, supports this hypothesis. And the right side of **Result 2.2** concerns the probability of the very same hypothesis, but this time only given the evidence of a single data point—I am a BB—and this particular data point anyway has questionable relevance to the hypothesis since it's not specified that I am a brain with evidence E. So, **Result 2.2**, which tells us that the right side is larger than

[31] In fact, **Lemma** is equivalent to **Result 2.1**, as is easily seen by examining the Appendix B proof.

[32] It also holds if we interpret H_E to mean that the proportion of brains with E that are BBs is above threshold *x*, where *x* can be any threshold between 0 and 1. We initially defined H_E to mean the "vast" majority of brains with E are BBs, which boosted the plausibility of **Defeat**, but **Lemma** will follow from **CI** regardless of where the threshold is.

the left side, seems to fly in the face of the most intuitive and ordinary standards of scientific reasoning. If **CI** is true, the inductive support relation is not what we think it is.[33]

We deny **CI**. We are proposing an alternative anti-skeptical position. The main argument of this paper has shown how we could happily endorse an ordinary, face value interpretation of any empirical evidence suggesting that we live in a BBU. We say the skeptical conclusion that *I'm* a BB would still not follow or even enjoy any support, not without further substantive premises such as **CI**. Without **CI** or something like it, there is no reason to be worried by a scientific theory that says there are zillions of BBs. Since we are comfortable rejecting **CI**, which has turned out to have odd anti-scientific implications, we are comfortable rejecting step 2 of the two-step argument.

In the end, the good news for everyone is at least that, whether **CI** is true or false, you can be sure there is no Boltzmannian threat of skepticism. If **CI** is true, then it would follow that we don't have good evidence that makes a BBU cosmology likely (Carroll's own view). And if **CI** is false, you can happily believe there are zillions of BBs without worrying at all that you are a BB (our own proposal). We like our proposal better because it requires no revisionary interpretations of our ordinary standards of scientific reasoning.

*c. Second Explanation, Continued: Diagnosing the Error in **CI***

Why does **CI**, an initially intuitive indifference principle, turn out to lead to such implausible results? The core thought motivating **CI** is an intuitive thought. The intuitive thought is that I must not treat myself as special: if I think a bunch of people have exactly the same evidence as I do, I cannot think that something is true of *me* without thinking it is also true of them. A more general version of the intuitive thought is this: the reasoning that I would apply to *myself* must be the same as the reasoning I would apply to *a randomly chosen evidential duplicate of myself*.[34] This is usually sensible advice. But not always.

---

[33] We take the particular example described in the text to be a decisive reason to reject **Result 2.2** (and so a decisive reason to reject **CI**). A more general argument against **Result 2.2** could take the following form. We could plausibly argue that, as we consider different bodies of evidence that are compatible with the conjunct OO, $Pr(H_E|E\&OO)$ can come arbitrarily close to 1, but $Pr(H_E|BB)$ *cannot* come arbitrarily close to 1—there is some limit *t* that is less than 1, and which $Pr(H_E|BB)$ cannot go above.

[34] As we mentioned earlier (fn.24), the advocates of **CI** motivate it by describing particular thought experiments involving "indistinguishable" duplicates and inviting us to have the intuition that subjects in any of those centers should be indifferent among them. Here we are trying to articulate the general version of that intuition.

It is bad advice when it comes to certain conditional probabilities. Sometimes, conditionalizing on a hypothesis about myself should give a very different result from conditionalizing on the corresponding hypothesis about a random duplicate of myself. And sometimes, the conditional probability that I am thus-and-so is very different from the conditional probability that a random duplicate of me is thus-and-so. We'll now illustrate how these things can come apart, and intuitively so. It's important to us to show they *intuitively* come apart because our goal now is to undermine the intuitiveness of **CI** and of **CI**'s motivation. For a first illustration, consider again this assumption that **Results 1** and **2** both relied on:

Defeat (partial statement): $Pr(H_E|BB) = Pr(H_E|BB\&E)$.

Notice, the claim is that E provides no evidence for H_E supposing *I* am a Boltzmann Brain. (Don't forget that "BB" abbreviates "*I'm a* BB" in our probability statements.) But E, of course, *can* provide evidence for H_E on the supposition that one of my duplicates, randomly selected, is a Boltzmann Brain. Why? Because supposing I am a BB, it's guaranteed that E describes a completely randomly generated experience, and so E has zero correlation with a hypothesis like H_E about how the universe is populated. (We've continued to assume the phenomenal view of evidence here, since **CI**'s plausibility depends on it, as we explained earlier.) But supposing only that a randomly selected duplicate is a BB, it's still a live possibility that I'm a reliable OO rather than a BB, and that will introduce some correlation between E and H_E. So, in this example, something that's true (**Defeat** is true) becomes false when we replace mention of me with mention of a random brain (it's not true that $Pr(H_E|$a randomly chosen duplicate of my brain is a BB$) = Pr(H_E|$E & a randomly chosen duplicate of my brain is a BB), for the reason we just gave). Thus, this is a case, a very clear and intuitive case, that does not behave in the way required by **CI**'s motivating thought that I should reason about myself exactly as I would reason about a randomly selected mind amongst my evidential duplicates. Thinking *I* am a Boltzmann Brain can have a special impact on the way I interpret my evidence—an impact that is not made by thinking *a randomly selected duplicate of mine* is a Boltzmann Brain.

For a second illustration, consider Pr(E|BB), a value which played a key role in our main argument, in section 2, against Boltzmann Brain skepticism. We argued that Pr(E|BB) was miniscule. Again, supposing *I'm* a BB, any experience I have is as likely as any other, which

makes the probability of having any particular experience vanishingly small, *including* an experience of scientists talking about Boltzmann Brains while apples rot. In contrast, supposing just that a random duplicate of mine is a BB (which leaves open the possibility that I'm an OO), it's not the case that any experience is as likely as any other.

What this example again illustrates is that "I" is not swappable with "a random duplicate of mine" in these contexts. The motivating intuition behind **CI** is giving bad advice again. And **CI** really is committed to giving this bad advice because the probability calculus entails[35] that Pr(E|BB) equals Pr(E|A random duplicate of mine is a BB) given the following assumptions, which **CI** is committed to:

(1) Pr(a random evidential duplicate of mine is a BB) = Pr(BB).

(2) Pr(BB|E) = Pr(a random evidential duplicate of mine is a BB|E).[36]

It seems to us, then, that in the end, not only does **CI** have absurd anti-scientific implications, but the motivating "I'm-not-special" idea that made **CI** initially intuitive does not really leave it ultimately intuitive after we've taken care to distinguish two very different kinds of thought. We must distinguish a thought about myself from the corresponding thought about a random brain like mine.

Finally, let's step back from the details of all the arguments we've now covered, and just pause to reflect again on the earlier point that our empirical evidence has egocentric content (as it must in order for us to make practical and theoretical inferences). While the most dedicated

---

[35] In general, Pr($X \mid Y$) = Pr($X \mid Z$) is entailed by the conjunction of Pr(Z) = Pr(Y) and Pr($Y \mid X$) = Pr($Z \mid X$).

[36] We'll sketch **CI**'s commitment to these, given an arbitrary possible world *w*. (When you are uncertain which of finitely many possible worlds is actual, your probabilities can be determined using the law of total probability in the usual way. But when there are infinitely many live possible worlds, **CI** must be strengthened slightly in order to get the intended verdicts of indifference. See Weatherson (2005, sec. 2) for the sort of strengthening of **CI** that he thinks (as does Builes (forthcoming, fn.24)) would naturally be accepted by **CI**'s adherents.)

In a centered worlds framework, given any (uncentered) possible world *w,* the probability that I'm a BB is the probability of the set *S* of all centers in *w* that are BB centers—everyone can agree to that. But because **CI** says every center in *w* is, apriori, equally probable, it will also say that the probability of *S* is equal to the probability that a randomly chosen center is in *S*. And the probability a randomly chosen center is in *S* is equal to the probability that a randomly chosen evidential duplicate of me is in *S.* (This is because every center is, apriori, a live possibility for being an evidential duplicate of me, so randomly choosing among my evidential duplicates is equivalent to randomly choosing among all centers). Thus, **CI** makes (1) true.

To see that **CI** will also make (2) true, notice that (2) will be true if Pr(E&BB) = Pr(E & a random evidential duplicate of me is a BB). The first value, Pr(E&BB), is the probability of the intersection of *S* with the set of centers that make E true. **CI** will say (as it does for every set) that the probability of this intersection is the probability that a randomly chosen center is in this intersection. (An opponent of **CI** can deny this, and our argument from section 2 justifies doing exactly that. We say that the probability E belongs to a BB can be lower than the probability a randomly chosen center with E belongs to an BB.) And **CI** will also say that the second value, Pr(E & a random evidential duplicate of me is a BB), is the probability that a randomly chosen center will be in that same intersection.

adherents of **CI** may not share our inclinations here, we trust we are not the only ones who'll find that the egocentricity of evidence casts some real doubt on the idea that I should always think of myself as I'd think of a randomly selected evidential duplicate. After all, my evidence is about *me*—not about a randomly selected duplicate of me. Perhaps, then, it shouldn't be that surprising that my evidence *can*, on special occasions, lead me to think something is true of *me*, but not true of most of my evidential duplicates.

4. Conclusion: the Empirical Case for Skepticism

We've framed this paper as a discussion of a skeptical worry posed by a promising scientific theory. As epistemologists, our concern was that, after all these centuries fighting the (traditional) skeptic, we may end up defeated by a new variety—one that comes armed with science. But many cosmologists, as we've seen, approach the issue quite differently from us: they are not worried by such nonsense as the possibility that they are floating solitary brains in space. Rather, they take the fact that an *otherwise* promising scientific theory leads to a skeptical worry to be a reductio against it. To the anxious epistemologists we say: do not fear the Boltzmann Brains—your anti-skepticism can remain intact. And to the pragmatic cosmologists we say: do not dismiss an otherwise promising scientific theory on the grounds that it has skeptical consequences—for it does not.

Although we don't believe we've received any empirical evidence for a skeptical hypothesis *yet*, is it possible that some day we will?

Our main positive argument, the argument of section 2, does not show that there is no possible empirical evidence that could support a skeptical conclusion. We do think certain very odd kinds of empirical evidence would qualify, such as Kotzen's (2021, p.26) example of experiencing a digital ticker tape running across your visual field with some message like, "You're in the Matrix but hold tight because we're trying to get you out!". But what Kotzen's case again shows us is that everything depends on the likelihoods.

In the example of the ticker tape, the likelihoods lead to skepticism. Your experience is intuitively good empirical evidence for skepticism because the experience is intuitively much likelier if you're in the Matrix than if you're an OO.

In some cases, it is not so obvious what the likelihoods are, and so the seriousness of the threat of skepticism should not be taken to be so obvious either. This is how we view the

skeptical argument that, given our empirical evidence, we almost certainly live in a simulation. For example, if Bostrom (2003, 2005) and Chalmers (2022, ch.5) are right, then we should be very worried by the empirical evidence that the vast majority of minds are computer simulations.[37] Hirsch (2018) describes a fictional case, a story about the discovery of a kind of brain emporium, which he suggests gives its characters empirical evidence they are envatted brains. Elga (2004, p.393) attributes a similar brain emporium story to Hartry Field and endorses the claim that it would have skeptical implications. The authors putting forward these skeptical arguments endorse **CI** or similar indifference reasoning, but, if our arguments of the last section are correct, indifference reasoning will not make for a plausible or an effective argument for skepticism from empirical evidence. The right way to evaluate this sort of empirical evidence (evidence that, at face value, suggests there is, or will be, a large number of envatted brains or simulated minds) and judge whether it confirms a skeptical conclusion about yourself is, as always, to check the likelihoods: is your experience more likely given the skeptical hypothesis than it is given the alternative? But we think these cases are very different from Kotzen's ticker tape case, where the intuitive likelihoods lead directly to skepticism. We have not made any arguments about what the values of the likelihoods are in these cases, and we leave it to others to properly examine the issue in detail. All we hope we have shown here is that the likelihoods hold the key to the strength or weakness of such a skeptical argument.

In the case of Boltzmann skepticism, we have argued for some specific claims about the likelihoods. Because most BBs live in a void and undergo incoherent experiences, we were able to argue that our empirical evidence is not likelier on the hypothesis that I'm a BB than on the hypothesis I'm an OO. That is why I'm not a Boltzmann Brain.[38]

[37] Officially, both Bostrom and Chalmers argue for disjunctive conclusions: "you are almost certainly a computer simulation *unless* … ." They each give some live alternative possibilities such as (a) obstacles prevent civilizations from reaching a sufficiently advanced technological stage, or (b) it is technologically difficult or impossible to run computer simulations of human minds, or (c) running simulations will always be taboo and successfully prohibited, or various other possibilities. But the question whether all of their alternatives are false seems to be an empirical question, not an apriori one. Bostrom also says it's partly for empirical reasons that the conditional itself is true: you're probably a simulation if none of his alternatives is true.

[38] Thanks to Zach Barnett, David Builes, Brian Cutter, Laura Callahan, Josh Dever, Harvey Lederman, John Pittard, Karl Schafer, Joel Velasco, two excellent referees, and audiences at Goethe University, Innland Norway University, Notre Dame, Texas Tech, and a meeting of the central APA.

Appendix A: Proof of Result 1

E: my total evidence (whatever it may be)
BB: I am a BB
H_E: the vast majority of minds with E are BBs

1. $Pr(BB|E\&\sim H_E) < Pr(BB|E\&H_E)$ [***premise:*** Lemma]

2. $Pr(H_E|E\&BB) < ½$ [***premise:*** Defeat]

3. $Pr(H_E|E\&BB) < Pr(\sim H_E|E\&BB)$ [7, negation law]

Now recall a general equation, which we introduced on p.5 and trivially follows from Bayes' theorem: the ratio of posteriors = the ratio of priors multiplied by the Bayes factor. The following is an instance of that equation, using the probability function that is conditional on E:

[[[—unfortunately google docs can't typeset the E subscript we intend to have on H here—]]]

4. $$\frac{Pr(H|E\&BB)}{Pr(\sim H|E\&BB)} = \frac{Pr(H|E)}{Pr(\sim H|E)} \times \frac{Pr(BB|E\&H)}{Pr(BB|E\&\sim H)}$$ [by Bayes' theorem]

(posteriors) (priors) (Bayes factor)

Rearranging that slightly gives us an equation that makes our conclusion easy to derive:

5. $$\frac{Pr(BB|E\&\sim H)}{Pr(BB|E\&H)} \times \frac{Pr(H|E\&BB)}{Pr(\sim H|E\&BB)} = \frac{Pr(H|E)}{Pr(\sim H|E)}$$ [algebra on 4]

6. $Pr(H_E|E) < Pr(\sim H_E|E)$. [1, 3 and 5, algebra]

7. $Pr(H_E|E) < ½$ [6, negation law]

Appendix B: Proof of Result 2

E: my total evidence (whatever it may be)
BB: I am a BB
H_E: the vast majority of minds with E are BBs

We'll prove our desired results by applying two familiar properties of Bayesian confirmation:

(i) $Pr(X|Y) > Pr(X)$ iff $Pr(Y|X) > Pr(Y)$[39]
(ii) $Pr(X|Y) > Pr(X)$ iff $Pr(X|Y) > Pr(X|{\sim}Y)$[40]

These hold for any X, Y, and any probability function. In particular, we'll be considering the probability function that is always conditioned on E.

1.	$Pr(BB|{\sim}H_E\&E) < Pr(BB|H_E\&E)$	[***premise:*** Lemma]
2.	$Pr(BB|E) < Pr(BB|H_E\&E)$	[1, ii]
3.	$Pr(H_E|E) < Pr(H_E|BB\&E)$	[2, i]
4.	$Pr(H_E|E\&{\sim}BB) < Pr(H_E|E\&BB)$	[3, ii]
5.	$Pr(H_E|E\&OO) < Pr(H_E|E\&BB)$	[4, replacing ~BB w/ OO]
6.	$Pr(H_E|E\&BB) = Pr(H_E|BB) < ½$	[***premise:*** Defeat]
7.	$Pr(H_E|E\&OO) < ½$	[5, 6, logic of =, <]

[39] $Pr(X|Y) > Pr(X)$
iff $Pr(X\&Y) > Pr(X)Pr(Y)$
iff $Pr(Y|X) > Pr(Y)$.

[40] $Pr(X|Y) > Pr(X)$
iff $Pr(X|Y) > Pr(X|Y)Pr(Y) + Pr(X|{\sim}Y)Pr({\sim}Y)$,
iff $Pr(X|Y) - Pr(X|Y)Pr(Y) > Pr(X|{\sim}Y)Pr({\sim}Y)$,
iff $Pr(X|Y)\,[1 - Pr(Y)] > Pr(X|{\sim}Y)[1 - Pr(Y)]$,
iff $Pr(X|Y) > Pr(X|{\sim}Y)$.

Bibliography

- Avni, Ron, 2023, "The Boltzmann Brains Puzzle", *Nous*
- Bostrom, Nick, 2003, "Are We Living in a Computer Simulation?", *Philosophical Quarterly*
- Bostrom, Nick, 2005, "The Simulation Argument: Reply to Weatherson", *Philosophical Quarterly*
- Builes, David, forthcoming, "Center Indifference and Skepticism", *Nous*
- Carroll, Sean, 2021, "Why Boltzmann Brains Are Bad", in *Current Controversies in Philosophy of Science*, eds. Dasgupta, Dotan, and Weslake, Routledge
- Chalmers, David, 2002, *Reality+*, Norton
- Comesana, Juan, 2020, *Being Rational and Being Right*, Oxford
- Crupi, Vincenzo, Branden Fitelson, and Katya Tentori (2008), "Probability, Confirmation, and the Conjunction Fallacy", *Thinking and Reasoning*
- Dyson, Lisa, Matthew Kleban, and Leonard Susskind, "Disturbing Implications of a Cosmological Constant", *Journal of High Energy Physics*
- Dogramaci, Sinan, 2020, "Does My Total Evidence Support that I'm a Boltzmann Brain?", *Philosophical Studies*
- Elga, Adam, 2004, "Defeating Dr. Evil with Self-Locating Belief", *Philosophy and Phenomenological Research*
- Hirsch, Eli, 2018, *Radical Skepticism and the Shadow of a Doubt*, Bloomsbury
- Kotzen, Matthew, 2012, "Dragging and Confirming", *Philosophical Review*
- Kotzen, Matthew, 2021, "What Follows from the Possibility of Boltzmann Brains?", in *Current Controversies in Philosophy of Science*, eds. Dasgupta, Dotan, and Weslake, Routledge
- Lewis, David, 1996, "Elusive Knowledge", *Australasian Journal of Philosophy*
- Miller, Brian, 2016, "How to Be a Bayesian Dogmatist", *Australasian Journal of Philosophy*
- Perry, John, 1979, "The Problem of the Essential Indexical", *Nous*
- Pryor, James, 2000, "The Skeptic and the Dogmatist", *Nous*
- Pryor, James, 2013, "Problems for Credulism", in *Seemings and Justification*, ed. Tucker, Oxford
- Quine, W. V., 1951, "Two Dogmas of Empiricism", *Philosophical Review*
- Saad, Bradford, "Lessons from the Void: What Boltzmann Brains Teach", *Analytic Philosophy*
- Schellenberg, Susanna, 2018, *The Unity of Perception*, Oxford
- Siegal, Susanna, 2021, "The Contents of Perception", in *Stanford Encyclopedia of Philosophy*
- Wallace, David, 2023, "A Bayesian Analysis of Self-Undermining Arguments in Physics", *Analysis*
- Weatherson, Brian, 2005, "Should We Respond to Evil with Indifference?", *Philosophy and Phenomenological Research*
- White, Roger, 2006, "Problems for Dogmatism", *Philosophical Studies*
- Williamson, Timothy, 2000, *Knowledge and Its Limits*, Oxford