Lessons from the Void: What Boltzmann Brains Teach

Runner-up for the 2022 Sanders Prize in Philosophy of Mind
Forthcoming in *Analytic Philosophy*

Bradford Saad (Utrecht University; Sentience Institute)

Abstract: Some physical theories predict that almost all brains in the universe are *Boltzmann brains*, i.e. short-lived disembodied brains that are accidentally assembled as a result of thermodynamic or quantum fluctuations. Physicists and philosophers of physics widely regard this proliferation as unacceptable, and so take its prediction as a basis for rejecting these theories. But the putatively unacceptable consequences of this prediction follow only given certain philosophical assumptions. This paper develops a strategy for shielding physical theorizing from the threat of Boltzmann brains. The strategy appeals to a form of phenomenal externalism about the physical basis of consciousness. Given that form of phenomenal externalism, the proliferation of Boltzmann brains turns out to be benign. While the strategy faces a psychophysical fine-tuning problem, it both alleviates cosmological fine-tuning concerns that attend physics-based solutions to Boltzmann brain problems and pays explanatory dividends in connection with time's arrow.

Keywords: consciousness; Boltzmann brains; phenomenal externalism; phenomenal internalism; interpretationism; cosmological fine-tuning; psychophysical fine-tuning; multiverse hypothesis; design hypothesis; axiarchic hypothesis; time's arrow; self-locating belief

Word count: 11,614 (main text)

Acknowledgements: For helpful feedback or discussion, I'm grateful to Brian Cutter, Hilary Greaves, Andrew Lee, Han Li, Andreas Mogensen, and participants in the 2022 Rutgers Philosophy of Religion Reading Group and participants in the 2022 Central European University summer seminar on The History and Philosophy of the Concepts of Scientific Law and Probability.

Lessons from the Void: What Boltzmann Brains Teach

1. Introduction

Some physical theories predict that almost all brains in the universe are *Boltzmann brains*,[1] i.e. short-lived disembodied brains that are accidentally assembled as a result of thermodynamic or quantum fluctuations.[2] Physicists and philosophers of physics widely regard this proliferation as unacceptable, and so take its prediction as a basis for rejecting these theories. Exactly what is supposed to be objectionable about such proliferation is a disputed matter, one that we'll have occasion to examine below.

For now, what's important is that predicting the proliferation of Boltzmann brains does not *by itself* put physical theories into conflict with any scientific data. These theories yield objectionable consequences only when their predicted preponderance of Boltzmann brains is *conjoined with substantive philosophical assumptions* about the epistemology of self-locating information and the distribution of consciousness. What's more is that some of the physical theories in question are otherwise appealing—appealing enough for physicists to turn to new physics to avoid the unwanted prediction. Proposed physics-based strategies for avoiding that prediction tend to introduce exotic posits. For example, one physics-based strategy posits a vacuum decay rate that would result in the destruction of our universe before Boltzmann brains become dominant.[3] Another such strategy posits a slow drift of physical 'constants' towards values that inhibit the creation of Boltzmann brains.[4]

This is a curious situation: scientists are substantially adjusting their physical theorizing in order to respect philosophical assumptions, assumptions that both elicit objectionable predictions from physical theories and concern matters of ongoing philosophical controversy. It is natural to wonder what it would take to extricate

---

[1] The label is due to Albrecht & Sorbo (2004).

[2] These theories include some versions of statistical mechanics, some multiverse theories, and some contemporary cosmological models of our universe, arguably including the current best-fit cosmological model (ΛCDM)—see Carroll (2020) for discussion and references. I'll discuss physical theories that proliferate Boltzmann brains in more detail in §§8-9.

[3] See Page (2008) and Boddy et al. (2013).

[4] See Carlip (2007). For other physics-based suggestions for avoiding Boltzmann brain production, see Boddy et al. (2016), Earman (2006), and Norton (2020).

physical theorizing from these philosophical shackles. In other words, what philosophical adjustments would allow physical theorizing to proceed without concern for whether a given physical theory proliferates Boltzmann brains? Some recent work bears on this question by developing shielding strategies that contest epistemological assumptions that figure in the derivation of objectionable consequences from Boltzmann brains.[5] This paper approaches the question from a different angle by developing a novel shielding strategy. My strategy is to reject assumptions from philosophy of mind that are used to derive objectionable consequences from theories that proliferate Boltzmann brains.

The key to my strategy is to embrace a certain form of *phenomenal externalism*, here understood as the thesis that worlds with the same laws as ours contain internal physical duplicates that differ phenomenally. More specifically, the key is to embrace a form of phenomenal externalism that attributes consciousness to ordinary brains and *zombifies* their Boltzmann counterparts, i.e. entails that they are unconscious. While zombification deprives Boltzmann brains of their worrying consequences, the story will not end here. In the wake of Boltzmann brain problems, analogous problems will emerge from Boltzmann creatures that are produced by the same processes that generate Boltzmann brains which duplicate ordinary observers in local—not just internal—physical respects. It will turn out that some forms of phenomenal externalism that suffice to shield physics from Boltzmann brain problems do not suffice to shield it from Boltzmann creature problems. But I will sketch a radically externalist theory that shields physics from both Boltzmann brain problems and Boltzmann creature problems. The theory is radically externalist in that it renders each subject's experiences directly sensitive to the entire state of the universe. After addressing some objections to the theory, I show how the theory bears on fine-tuning problems and proposed explanations of time's thermodynamic arrow. I will conclude by drawing morals about how Boltzmann brains and their ilk forge evidential connections between physics and philosophy of mind.

[5] See Carroll (2020), Dogramaci (2020), Dorr & Arntzenius (2017), Hartle & Srednicki (2007); cf. Kotzen (2020).

2. What's Bad about Boltzmann Brains?

Physical theories that proliferate Boltzmann brains arguably have three objectionable consequences: skepticism, cognitive instability, and an expectation of disorder that is not borne out in our experiences. This section will develop these *Boltzmannian objections* and pinpoint their philosophical sources. This will set the stage for §3, which will show how to answer the Boltzmannian objections in a single theoretical stroke.

Before proceeding, two terminological preliminaries are in order. First, it is standard to define 'Boltzmann brains' in a way that entails that they are conscious. It is also standard to assert that Boltzmann brains are predicted to (statistically) dominate the brains of ordinary observers on certain physical theories. However, the physical theories in question say nothing whatsoever about consciousness and so do not predict that there are any Boltzmann brains, if being a Boltzmann brain definitionally requires being conscious. At best, these theories predict Boltzmann brains on such a definition conditionally upon an implicit philosophical assumption about which physical systems are conscious. One goal of this paper is to bring such assumptions into the open and explore how they would need to be adjusted to shield physical theories. For this purpose, it will be useful to be able to talk about the systems proliferated by the relevant physical theories in a way that's neutral about whether those systems are conscious. That's why I have defined Boltzmann brains in a way that leaves open whether they are conscious.

Second, the physical theories of interest do not explicitly say anything about lumps of gray matter like the ones we find in our head. We know that they proliferate such configurations of matter because of a general feature of these theories: for any given internal physical state, they predict that almost all instances of that state will result from small fluctuations away from equilibrium.[6] One corollary is that these theories would proliferate internal physical states of conscious systems regardless of whether those systems have brains—if brainless alien organisms or computers can be conscious, Boltzmannian instances of their internal physical states will dominate the

[6] See, e.g., Carroll (2020: 10).

ordinary ones on these theories. Unfortunately, in what follows we cannot simply focus on brains and ignore the possibility of conscious but brainless physical systems: on pain of ignoring live hypotheses about multiple realizability, the latter must be taken into account in some of the Boltzmannian objections' statistical reasoning. On the bright side, this can be done by extending what I say about brains to brainless physical systems. Repeatedly making such extensions would be tedious. As a workaround, I propose to stretch the meaning of 'brain' to cover all sorts of internal physical hardware of conscious systems along with internal physical duplicates thereof (whether or not the duplicates are conscious). Likewise, I'll extend 'brain state' to the internal physical states of such systems.

2.1 Skepticism

The first objection against theories that proliferate Boltzmann brains is that they are unacceptable because they lead to skepticism. Here's a first pass formulation of the skeptical objection: if one of these theories is true, almost all observers with your experience are Boltzmann brains (in the stretched sense of 'brain') and therefore in skeptical situations—i.e. situations in which most of their ordinary beliefs are mistaken—in which case you're probably in such a situation. But it's improbable that you're in a skeptical situation. So, these theories are not true.

The path from these physical theories to skepticism is paved by two philosophical claims. One is *phenomenal internalism*, the popular but controversial thesis that, at least within worlds with the same laws as ours, internal physical states fix phenomenology. In other words, phenomenal internalism holds that it's nomically impossible for entities to differ phenomenally without differing in their brain states (in the stretched sense of 'brain state'). The second is the *self-locating indifference principle* that one should divide one's credence equally among hypotheses about one's location among those that are subjectively indistinguishable from one's own.[7]

We can now raise the specter of skepticism. These physical theories predict that almost all instances of any given brain state belong to Boltzmann brains. This, of

[7] See Bostrom (2002*a*: Ch. 10) and Elga (2004).

course, applies to the brain states, if such there be, that realize[8] the type of experience you are now having, i.e. the type encompassing exactly those experiences that are subjectively indistinguishable from your present one. Given phenomenal internalism, there is at least one such brain state. Therefore, combining these physical theories with phenomenal internalism yields the result that almost all realizers of the type of experience you are now having belong to Boltzmann brains.[9] But if almost all realizers of the type of experience you are now having belong to Boltzmann brains, then almost all instances of that type of experience also belong to Boltzmann brains.[10] So, given phenomenal internalism, these theories predict that almost all instances of the type of experience you are now having belong to Boltzmann brains. These brains are, of course, skeptically situated: they share your beliefs about being an embodied member of an evolved species on the surface of a planet; but they are mistaken on all these counts and more besides. Given the prediction that almost all instances of the type of experience you are now having belong to such Boltzmann brains, the self-locating indifference principle enjoins you to conclude that you are almost certainly a Boltzmann brain and therefore skeptically situated.

2.2 Instability

Another complaint against these theories is that they are unacceptable because they are cognitively unstable: accepting them undermines evidence for them.[11]

The worry can be developed as follows. Suppose that, on the basis of astronomical data, you came to accept a physical theory that proliferates Boltzmann

[8] I'll understand realization in terms of nomic sufficiency: for *x* to realize *y* is for *x* to entail *y*, given the laws of nature.

[9] At least if we set aside scruples about infinitary aggregation in the event that there are infinitely many such states.

[10] At any rate, this is so provided that we set aside bizarre ways in which realizers can conceivably numerically diverge from the experiences they realize. For instance, we need to exclude the possibility that each realizer of an orderly experience realizes infinitely many orderly experiences while each realizer of a disorderly experience realizes only finitely many disorderly experiences.

[11] See Carroll (2020), Crawford (2013), Chalmers (2018: fn25), Myrvold (2016), and North (2011); cf. Dogramaci (2020) and Barrett (1996). This objection is sometimes attributed to David Albert (2000). However, as others (Dogramaci, 2020; Chen, 2021) have noted, Albert does not discuss cognitive instability there, though he discusses threats of skeptical catastrophe and self-undermining (2000: 116). For a response to the instability objection, see Kotzen (2020).

brains. As we saw in §2.1, given phenomenal internalism, the theory will also predict that almost all experiences like yours belong to Boltzmann brains. Upon recognizing this, you would then need to apply the self-locating indifference principle and conclude that you are almost certainly a Boltzmann brain. But Boltzmann brains are in no position to do astronomy or access astronomical data. Thus, upon accepting that you are almost certainly a Boltzmann brain you should recognize that your putative astronomical grounds for this conclusion should not be trusted. Upon regaining confidence that you are not a Boltzmann brain, you should also regain trust in those grounds. Thus, given phenomenal internalism and the self-locating indifference principle, these theories put you in a cognitively unstable situation: a situation in which you should take yourself to have data that shows that you are almost certainly a Boltzmann brain just to the extent that you are confident that you are not a Boltzmann brain.

2.3 Disorder

The remaining objection is: on physical theories that predict Boltzmannian domination, almost all experiences would belong to Boltzmann brains and be disorderly. In that case, on these theories, I should expect with near certainty to have a disorderly experience rather than an orderly one. Therefore, my orderly experience provides strong evidence against these theories.[12]

In its present form, the objection is underdeveloped. It invokes a notion of experiential disorder that is somewhat intuitive but imprecise. However it is precisified, one wonders why we should expect almost all experiences to be disorderly. Unfortunately, to my knowledge, proponents of the objection have not precisified the operative notion of disorder; nor have they unpacked the objection significantly beyond its present form. I do not have any suggestions for how to precisify experiential disorder. So I will simply assume the availability of some such notion and offer my best guess at what proponents of the objection have in mind.

[12] For objections in the vicinity, see Carlip (2007), Chalmers (2022: Ch. 24), Kotzen (2020), Page (2008), Norton (2020). For some doubts about this objection, see Crawford (2013).

To start, we can picture the realization of experiences in Boltzmann brains as follows. A theory predicts a vast number of fluctuations, a small portion of which yield a vast number of Boltzmann brains. Whenever a Boltzmann brain fluctuates into existence, we can think of the fluctuation that produces it as sampling from the space of brain states that can exist in worlds with laws like ours. In accordance with phenomenal internalism, the objection requires that some fluctuations would yield conscious Boltzmann brains. We can think of these fluctuations in like manner as sampling from the space of brain states that realize experiences. In addition, we can think of brain states that realize experiences as sampling from a space of experiences (the space that can be realized by brain states in worlds with the same laws as ours). Now, to determine how likely a randomly selected experience of a Boltzmann brain is to be orderly, we can ask, how likely is such a sampling sequence to yield an orderly experience rather than a disorderly one? The answer to that question depends on several factors:

- the prevalence of orderly experiences within the noted space of experiences,
- the bias (if any) among brain states that realize experiences toward realizing orderly experiences, and
- the bias (if any) among posited fluctuations that result in brain states that realize experiences toward brain states that realize orderly experiences.

To illustrate, suppose that disorderly experiences dominate the space of experiences that can be realized by brain states. In that case, so long as brain state realizers of experiences are not strongly biased toward realizing orderly experiences, a given randomly selected brain state from the space of those that realize experiences will almost certainly realize a disorderly experience. Supposing that disorderly experiences dominate the noted space and that such a realization bias is absent, how likely a randomly selected experience among those the result from fluctuation-induced brain states will depend on fluctuation bias: if fluctuations that produce experiences via brain states are not strongly biased toward brain states that realize orderly experiences, then almost all experiences produced by such fluctuations will be disorderly.[13]

I suggest that the objection be construed as embracing these suppositions. The initial supposition (that disorderly experiences dominate) can be motivated by noting that, intuitively, there seem to be vastly more disorderly ways of combining aspects of experience than there are orderly ones. Moreover, absent positive reasons to the contrary, it seems reasonable to assume that brain states that realize experiences and fluctuations that yield the brain states of Boltzmann brains are not strongly biased toward orderly experiences, i.e. not biased enough to prevent it from being almost certain—conditional on disorderly experiences dominating—that an arbitrary fluctuation yielding a brain state that realizes an experience will yield a brain state that realizes a disorderly experience.[14]

In this fashion, the objector can show that a given conscious Boltzmann brain is almost certainly having a disorderly experience rather than an orderly one. From here, she can reason, if Boltzmann brains dominate ordinary observers, brains with disorderly experiences dominate brains with orderly experiences. This does not quite show that I should expect *ex ante* to have a disorderly experience on theories that predict that Boltzmann brains dominate ordinary observers. But that conclusion can be reached through two further assumptions. The first, which accords with phenomenal internalism, is that a brain state realizes my experience. The second is that, conditional on a brain state realizing my experience and on a theory that predicts a certain proportion of experiences realized by brain states will be disorderly, my credence that my experience will be disorderly should match the predicted proportion. It follows, then, that on theories predicting that Boltzmann brains dominate ordinary observers, I should be almost certain that my experience will be disorderly. On the rival hypothesis

[13] As before, I am setting aside bizarre ways in which realizers can conceivably numerically diverge from the experiences they realize.

[14] An additional motivation that the objector might cite is that fluctuations are more likely to produce *entropically*-disorderly brain states than they are to produce entropically-orderly ones; so, given a positive correlation between the entropic-disorder of brain states and the disorder of experiences (however exactly that notion is cashed out), fluctuations are biased *toward* brain states that realize disorderly experiences.

that ordinary observers dominate (or at least are not dominated by Boltzmann brains), it is presumably much less likely that my experiences will be disorderly.

Thus we have a kind of crucial experiment: if my experience is orderly, that is strong evidence in favor of the hypothesis that ordinary observers dominate over physical theories on which Boltzmann brains dominate. Since my experience is in fact orderly, I therefore have strong evidence against theories that predict that Boltzmann brains dominate ordinary observers. Or so says the objection according to my best guess at what it is supposed to be.

One strategy for shielding physical theories from this objection embraces the *order bias hypothesis* that fluctuations that produce conscious Boltzmann brains are somehow biased toward producing brain states that realize orderly experiences, biased enough for the fact that my experience is orderly to not provide much evidence against theories that predict Boltzmannian domination. The bias in question might trace to a fluctuation bias, a realization bias, or a bias toward order in the space of experiences. As far as it goes, I think this is a promising strategy for dealing with the disorder objection. However, taken by itself, it has an unfortunate limitation: it cannot be naturally extended to shield theories that predict Boltzmannian domination from the charges that they lead to skepticism and instability. For the skeptical and instability objections can be raised while considering Boltzmann brains that only have orderly experiences like the one you are now having. In §5, I will explore a strategy that overcomes this limitation by deriving an order bias from a theory that also answers the skeptical and instability objections.

3. The Zombie Shield

The foregoing objections have a common thread: phenomenal internalism. Whereas the objections from skepticism and instability directly appeal to phenomenal internalism, the disorder objection makes two assumptions that will be accepted by virtually all phenomenal internalists and rejected by virtually all its opponents: namely that some fluctuations would produce conscious Boltzmann brains and that a brain state realizes my experience. Rejecting phenomenal internalism is therefore a natural move for those who wish to get on with physical theorizing without having to take care to steer clear of Boltzmannian domination.

This move does not quite work, as the objections can be recast in terms of (hitherto unformulated) theses that are weaker than phenomenal internalism. For instance, all of the objections can be run just as well using the thesis that, of nomic necessity, experience is preserved under the duplication of internal physical states *in systems that have the extrinsic features shared by ordinary observers and their Boltzmann brain counterparts.* Similarly, all of the objections might be run just as well using the thesis that, of nomic necessity, internal physical duplicates have somewhat similar experiences. It is unclear what the weakest thesis entailed by phenomenal internalism is that can be used to run all these objections. I will not try to resolve that issue here. Instead, I will show how to answer these objections with a phenomenal externalist thesis that is clearly strong enough to block the objections. That thesis is:

> **Zombified Brains:** there is some or other extrinsic physical requirement for consciousness that no Boltzmann brain would satisfy; so no Boltzmann brain is conscious.

Zombified Brains makes quick work of all three objections.

First, the objection from skepticism: Zombified Brains entails that no conscious Boltzmann brain is skeptically situated. Upon recognizing this, I can reason that since I am conscious and no Boltzmann brain is conscious, I am not a skeptically situated Boltzmann brain. This reasoning goes through just as well on the supposition that a theory that predicts Boltzmannian domination is true. Similarly, if we accept Zombified Brains, instability ceases to threaten theories that predict Boltzmannian domination: given Zombified Brains and that I'm conscious, I could accept a theory that predicts Boltzmannian domination on the basis of evidence without then being under pressure to conclude that I am a Boltzmann brain and, in turn, that I do not have that evidence. Finally, the objection from disorder requires Boltzmann brains to have disorderly experiences on theories that predict Boltzmannian domination; but, given Zombified

Brains, on these theories Boltzmann brains do not have any experiences, much less disorderly ones.

In this fashion, Zombified Brains renders hordes of Boltzmann brains benign. Hence, by embracing Zombified Brains, we can shield physical theories that predict Boltzmannian domination from the objectionable consequences that are supposed to follow. However, simply appealing to Zombified Brains and leaving it at that would not make for a satisfactory shielding strategy. That would invite the charge of ad hocery. Still worse, some philosophers would reject Zombified Brains because they think that the truth of phenomenal internalism can be established from the armchair.

Before motivating Zombified Brains, it's worth explaining why the thought that phenomenal internalism can be established from the armchair is misguided. The thought is supposed to be the deliverance of intuition. But we can conceive both of worlds in which phenomenal internalism is true and of worlds in which phenomenal internalism is false. Neither world need involve anything that we are a priori justified in regarding as non-actual—for instance, neither world need involve our being skeptically situated. Whether phenomenal internalism is true is therefore not something we can establish a priori. Nor does a priori reflection, when aided by introspection, come down in favor of phenomenal internalism. If anything, the naive realist intuition that features of the environment (shapes, for example) help constitute veridical experiences tells against phenomenal internalism. Nor does the fact (arguably recognized by commonsense) that differences in the brain make for differences in experience clinch the case for phenomenal internalism: one can reject phenomenal internalism while allowing for such internal difference makers.[15]

One motivation for Zombified Brains is that it avoids a puzzle that *hypothetical* Boltzmann brains raise for phenomenal internalism and other rival hypotheses independently of theories that predict Boltzmannian domination. To illustrate the puzzle, let's consider a nomically possible scenario in which a Boltzmann brain is an internal physical duplicate of you as you have an experience as of a blue triangle.

---

[15]For arguments in the same spirit, see Dalbey & Saad (2022), Pautz (2014), and Saad (2019*a*).

Phenomenal internalism predicts that your disembodied counterpart also has an experience as of a blue triangle. For phenomenal internalists, the puzzle is that of identifying a systematic psychophysical principle that explains why the Boltzmann brain in this case has an experience as of a blue triangle. The trouble is that we would expect a systematic principle to operate on a discriminating physical connection, a connection between the Boltzmann brain and the property it experiences (that of being a blue triangle, according to phenomenal internalism) but not between the brain and properties it does not experience. Yet there seems not to be any such physical connection between the Boltzmann brain and the property of being a blue triangle. For instance, its brain state does not produce behavior that is appropriate, given an experience as of a blue triangle. And given the many ways it could be embodied in nearby worlds, there is no fact of the matter about how it would behave if it were embodied; much less is there such a fact of the matter that connects it with the property of being a blue triangle. It does not perceive that property. Nor is it a member of a community in which that brain state normally occurs in response to a blue triangle.

The puzzle arises in the same form for theorists who reject phenomenal internalism but accept its prediction about what your Boltzmann brain counterpart experiences. The puzzle arises in a different form for theorists who claim that the Boltzmann brain counterpart's experience differs from yours. For them, the puzzle is that of identifying a systematic psychophysical principle that explains both why you experience blue triangularity and why your Boltzmann brain counterpart experiences a certain other property. On this form of phenomenal externalism, we'd expect such a systematic principle to operate on a discriminating physical connection that differentially relates you and your Boltzmann brain counterpart to the properties you respectively experience. But it is very hard to see what that connection might be: the discriminating physical connections (for instance, spatial and causal relations) you bear to blue triangularity but not other colors or shapes seem not to connect the Boltzmann brain to anything. And it would seem that any discriminating physical connection that relates the Boltzmann brain to a property would also relate you to that property.

By embracing Zombified Brains, we can avoid this puzzle. Since Zombified Brains entails that your Boltzmann brain counterpart is unconscious, it does not require a systematic psychophysical principle that applies to the Boltzmann brain counterpart or a discriminating connection between it and the properties it experiences. As it stands, this response is incomplete: if we cannot at least gesture at a systematic psychophysical principle that would, on Zombified Brains, explain why you have an experience as of a blue triangle and the Boltzmann brain is unconscious, then we have reason to doubt that this proposal will ultimately be better off than the phenomenal internalist and phenomenal externalist views for which we found no solution. This explanatory demand can be met by appealing to:

> **Simple tracking theory**: It's nomically necessary that a subject has a sensory experience as of *F* if and only if she is in an internal physical state that:
> (1) plays a suitable functional role, and
> (2) tracks *F*,
> where *x* tracks *F* iff *x* would covary with *F* under optimal conditions.

Different versions of this theory (or, if you prefer, theory schema) spell out the key notions of the theory in different ways.[16] For instance, some versions define the operative functional role in terms of accessibility to a global work space; others define it in terms of being in a state that is maximal with respect to the measure of integrated information. Some versions define covariation causally or perceptually; others define it modally or explanatorily. And some versions define optimal conditions in evolutionary terms—for example as the conditions under which a state is selected for; others define it in terms of statistically normal viewing conditions or the absence of interfering factors. We need not decide between these versions here.

What is important for our purposes is that this theory gestures at a systematic psychophysical principle that explains why you have the experience as of a blue triangle and why your Boltzmann brain counterpart has no experience at all: you have that experience because you are in an internal physical state that plays a suitable functional

[16] For physicalist theories along these lines, see Dretske (1995) and Tye (1995). For an overview of different ways of developing the theory and references, see Dalbey & Saad (2022).

role while tracking blue triangularity. And your Boltzmann brain counterpart is unconscious because its internal states do not track anything. Thus, the simple tracking theory provides a sort of proof of concept that Zombified Brains is compatible with a systematic account of experience. The simple tracking theory thereby shows that we can use Zombified Brains to resolve the puzzle without having to relinquish aspirations for such an account.

To sum up, Zombified Brains is a motivated thesis in philosophy of mind. By embracing it, proponents of physical theories that predict Boltzmannian domination can avoid the dire consequences that are alleged to follow from that prediction. I now turn to problems for these theories that lie in the wake of those posed by Boltzmann brains.

4. Boltzmannian Revenge

Physical theories that predict that Boltzmann brains dominate ordinary observers generally also predict that ordinary observers are dominated by *Boltzmann creatures.* A Boltzmann creature is a system with a brain and functioning body that inhabits a *Boltzmann bubble*, i.e. a configuration of matter that, like a Boltzmann brain, accidentally results from thermodynamic or quantum fluctuations.[17] This suggests a fallback strategy for those who, when confronted with Zombified Brains, abandon the objections to these theories from skepticism, instability, and disorder: run the objections using Boltzmann creatures rather than Boltzmann brains.

This strategy is subject to two important constraints. First, the strategy needs to appeal to Boltzmann bubbles that are small enough to generate the objections. On theories that predict them, Boltzmann bubbles become increasingly less common as they become larger—this is because larger bubbles would require fluctuations that are less probable. The objections under consideration essentially invoke Boltzmannian posits that dominate ordinary observers. Thus, one way to run afoul of the small-enough-bubble constraint is to object to a theory by invoking Boltzmann bubbles that do not contain enough Boltzmann creatures to dominate ordinary observers on that

[17]See Carroll (2020: 2020: 10) and Chen (2021); cf. Bostrom (2002: §IV) and Dainton (2012: §6).

theory. The danger of violating this constraint in this fashion may not come to much in practice, as theories on which Boltzmann brains dominate are also theories on which much larger Boltzmannian entities such as Boltzmann galaxies dominate. But the constraint may be violated in other ways. To illustrate, suppose that a theory predicts that Boltzmannian entities the size of the known universe dominate ordinary entities of that size. And then consider the objections from skepticism, instability, and disorder couched in terms of these enormous Boltzmann universes, i.e. systems that are physically like the known universe but which are fluctuations. These objections would fail. For the bubbles they invoke would be populated by ordinary observers rather than skeptically situated ones. Hence, they would not generate the skeptical or instability worries. And evolutionary processes would presumably prevail within such bubbles, leading them to contain more observers with orderly experiences than with disorderly experiences.

The second constraint is that the strategy cannot invoke Boltzmann bubbles that are too small, lest it succumb to slight extensions of the Zombified Brain response. For instance, it won't work to run the objections in terms of bubbles that consist of embodied brains whose ordinary environments extend no more than two meters in each direction. While these bubbles contain skeptically situated creatures, those creatures do not bear any of the standard candidates for tracking relations to features of their environment. Consequently, the simple tracking theory zombifies such Boltzmann creatures no less than it zombifies Boltzmann brains and so disarms objections from the former just as easily as it disarms objections from the latter.

For several reasons, satisfying both of these constraints is not straightforward. One technical obstacle is that of determining the maximal bubble size that ensures that Boltzmann creatures (of a suitable sort) will dominate ordinary observers, given both a physical theory and a reasonable operationalized precisification of 'ordinary'.[18] Another formidable obstacle is that of determining how large bubbles need to be in order to avoid succumbing to slight extensions of Zombified Brains. To illustrate, suppose

[18] See De Simone et al. (2010) for relevant discussion.

Zombified Brains is true because the simple tracking theory is true. Then how large bubbles need to be to overcome slight extensions of Zombified Brains reduces to the question of how large bubbles need to be in order for Boltzmann creatures they contain not to be zombified by the simple tracking theory. But that question turns on what account of tracking is plugged into the simple tracking theory. For instance, an account that construes tracking as a matter of perceptual covariation in the absence of interference might zombify only relatively small bubbles, ones just large enough for subjects to sustain perceptual contact with features of their environment. In contrast, if tracking is evolutionarily defined in terms of perception and conditions under which a internal states underpinning perception were selected for, then bubbles might need to span the surface of a planet and reach millions of years into the evolutionary past.

One strategy for preempting creature-based Boltzmannian objections would opt for a version of the simple tracking theory that is especially adept at zombifying large bubbles. The just suggested evolutionary version of the simple tracking theory might work for this purpose.[19] That simple tracking theory would at least block objections to physical theories, if such there be, that only predict Boltzmannian domination by creatures in bubbles that are too small to house the evolutionary processes that eventuate in conscious observers. For the evolutionary simple tracking theory would zombify all such creatures, in which case the objections would not get off the ground. However, I do not know of any physical theories that plausibly fit this description, as physical theories that predict Boltzmannian domination of any sort generally predict it courtesy of creatures in large bubbles. Whether the evolutionary simple tracking theory can shield this sort of physical theory from the objections is less clear cut. Let us consider what this theoretical package says about each objection in turn.

[19] Cf. Dretske (1995). I note in passing that *avoiding* the zombification of (what are in effect) Boltzmann creatures has been advanced as a problem for phenomenal externalist theories—the problem is often put in terms of Davidson's (1987) Swampman. As Pautz (2014: 161-2) notes, proponents of phenomenal externalism have been oddly concessive about the force of the problem. This section can be seen as turning the dialectic on its head by identifying a theoretical advantage that accrues to some phenomenal externalist theories by way of their zombifying powers.

- Plausibly, the evolutionary simple tracking theory answers the disorder objection against such physical theories: if the only bubbles with conscious inhabitants are large enough to contain evolutionary processes that create conscious creatures, then observers with orderly experiences would presumably outnumber observers with disorderly experiences.
- This theoretical package countenances conscious observers in relatively large bubbles but not in small ones. As a result, it does not obviously pose a skeptical threat. Granted, some observers in these bubbles would have many false beliefs about the astronomical happenings beyond their bubble. Even so, the everyday beliefs of observers in tracking-conducive bubbles might well qualify as knowledge, as what conduces to tracking may also conduce to knowledge.[20] While this hypothesis has a modicum of plausibility, it is not obviously true or easily verified. So, on this theoretical package, the skeptical objection is inconclusive.
- It seems doubtful that the astronomical data available to observers in bubbles would be trustworthy. *A fortiori*, it is doubtful that the astronomical data available to such observers that supports the physical theories in question would be trustworthy. Conditional on almost all observers being in bubbles, it is therefore doubtful that we have trustworthy astronomical data of the sort needed to support physical theories that predict Boltzmannian domination. So the instability objection poses a residual threat to the suggested theoretical package.

To sum up, the simple tracking theory shields physical theories from Boltzmann brains, but not from revenge at the hands of Boltzmann creatures. Supplementing the simple tracking theory with an evolutionary account of tracking held some promise for physical theories as a shield against Boltzmann creatures, but turned out to be inconclusively protective. In the next section I will explore a different shielding strategy, one designed to block not only the three brain-based Boltzmannian objections but also their creaturely cousins.

5. A Global Solution

I will now show how physical theories that predict Boltzmannian domination can be shielded from both Boltzmann brains and Boltzmann creatures. The strategy forgoes the simple tracking theory for a more radical form of phenomenal externalism, which I'll call *global interpretationism*.[21] According to global interpretationism, the distribution

[20] Cf. Nozick (1981).

[21] Global interpretationism is inspired by Lewis's interpretationism about non-phenomenal mental states (Lewis, 1974; 1994). Roughly, on Lewis's view, believing that p is a matter of being in a state that plays the belief-role while being assigned p on the best interpretation, where the best interpretation is the one that

of experience is fixed by a global interpretation principle. This principle takes the entire world as a minimal input and outputs whichever distribution is assigned by the best interpretation of that world according to a certain measure. More specifically, I will appeal to a version of global interpretationism, namely *K-max interpretationism*. On it, the operative measure ranks interpretations at worlds by giving positive weight to the interpretations' simplicity and to how much knowledge the interpretations entail at the world in question.[22]

K-max interpretationism does not fully specify the operative global interpretation principle. For instance, knowledge and simplicity are left at an intuitive level—a more fleshed out version of K-max interpretationism would need to tell us which precisifications of these notions figure in the principle. K-max interpretationism is also neutral on what sort of principle the global interpretation principle is. The principle could be construed reductively as an identification scheme on which having an experience as of *F* just is being in a state that is assigned *F* by the K-max global interpretation principle. Alternatively, the principle could be construed as a grounding law that specifies conditions under which states ground experiences that are irreducible to those states. Finally, the principle could be construed as a fundamental psychophysical law that specifies conditions under which experiences are caused. While I elsewhere (Saad, 2022) argue in favor of the latter construal over rivals, I will here remain neutral on the metaphysical status of the principle to which K-max interpretationism appeals.

Having introduced K-max interpretationism, I will now explain how it can shield physical theories from Boltzmannian objections. To start, consider internalist interpretations, i.e. ones that assign internal physical duplicates the same experiences. On physical theories that predict Boltzmannian domination, these interpretations place

---

strikes an optimal balance of various desiderata such as assigning agents beliefs and desires that are, given their evidence rational and rationalizing of their behavior.

[22] The view is so called because its distinctive claim is that experiences are assigned so as to maximize knowledge, provided this does not detract too much from simplicity. For knowledge-based interpretationist views of metasemantics, see Williamson (2000: 267; 2007: Ch. 8) and Cappelen & Dever (2021). For interpretationist views of consciousness that are not knowledge based, see Cutter & Crummett (forthcoming) and Pautz (2021).

almost all observers in skeptical scenarios. Given that a preponderance of skeptically situated observers detracts from knowledge on the operative measure, such interpretations will score poorly on that measure by the lights of those theories. Even if a preponderance of skeptically situated Boltzmann observers does not directly detract from knowledge on the operative measure, it will plausibly do so indirectly: if you are an ordinary observer and almost all phenomenal duplicates of you are skeptically situated, your situation is akin to someone in fake barn country looking at a real barn. Your true beliefs about the barn plausibly fall short of knowledge, due to their extraordinary luck; likewise, in the envisioned scenario of Boltzmannian domination, you plausibly lack knowledge about your ordinary environment that you would have if not for its embedding within a world that renders your beliefs extraordinarily lucky.[23] This reasoning does not depend on the self-locating indifference principle: just as fake barns imperil knowledge of barn observers who unwittingly inhabit fake barn country, so too do skeptically situated Boltzmann brains imperil knowledge of ordinary observers who live in blissful ignorance of, and disbelief in, such brains.

Next, consider phenomenal externalist interpretations—that is, interpretations that sometimes assign internal physical duplicates different experiences. As we have seen, mildly externalist forms of phenomenal externalism will not be enough to shield physical theories from the objections, as the objections can be recast in terms of Boltzmann creatures. Similarly, on physical theories that predict creatures in Boltzmann bubbles to vastly outnumber ordinary creatures, mildly externalist interpretations will lead to a preponderance of skeptically situated observers. As a result, these interpretations will score poorly on the K-max measure for the same reasons that phenomenal internalist interpretations do.

[23] Admittedly, intuitions about fake barn cases vary and are contested—see, e.g., Gendler & Hawthorne (2005). One might take this to undermine the proposed analogy. In response, I suggest that this reveals a constraint on how K-max interpretationism would need to precisify its operative notion of knowledge if it is to be enlisted along the lines envisioned in the main text. It should be borne in mind that whereas intuitions about knowledge may reveal more about our parochial concepts than the joint-carving categories in the vicinity (cf. Weatherson, 2003), we should expect psychophysical principles to be couched in terms of joint-carving categories rather than their less natural, parochial counterparts.

On the other hand, forms of phenomenal externalism that do not lead to a preponderance of skeptically situated observers will score better. For instance, consider an evolutionary version of the simple tracking theory on which the physical base of a given experience spans the surface of a planet and millions of years into the past. Plausibly, on that view, beliefs about the local environment formed on the basis of such experiences would constitute knowledge, knowledge that would be unthreatened by hordes of Boltzmann brains, per their zombification. As before, astronomical belief may fail to constitute knowledge in bubbles that are just large enough to be inhabited by conscious creatures. To the extent that such beliefs are intertwined with other non-astronomical beliefs, the failure of astronomical beliefs to qualify as knowledge may prevent some non-astronomical beliefs from so qualifying as well.[24]

While an evolutionary tracking interpretation would score better than internalist or mildly externalist interpretations, there is little reason to think that it would be the best interpretation on K-max interpretationism. In fact, there is reason to think that it would not be. The evolutionary tracking interpretation scores better than others because it requires a larger minimal size for Boltzmann bubbles containing conscious observers, thereby in effect ensuring that brains will be conscious only if they are in a position to know much about the world. Interpretations that require even larger minimal sizes for such bubbles can be expected to score even better.

Here's a recipe for constructing interpretations that zombify those creatures zombified by the noted evolutionary interpretation and larger ones besides: take any feature ψ that belongs only to larger bubbles than the smallest ones assigned experiences by the evolutionary tracking interpretation; then restrict the evolutionary tracking interpretation to bubbles of that size. Schematically:

[24] Cf. Chen (2021: §3.4).

> **Restricted Evolutionary Interpretation Schema**: It's nomically necessary that a subject has a sensory experience as of *F* if and only if she is in an internal physical state that:
>
> (1) has ψ
> (2) plays a suitable functional role, and
> (3) tracks *F*,
>
> where *x* tracks *F* iff *x* would covary with *F* under evolutionarily construed optimal conditions.

Plugging in features that belong only to sufficiently large bubbles will yield interpretations that assign experiences which would vindicate much of the astronomical knowledge we take ourselves to have, along with the knowledge that depends on it. Provided that some such feature is simple enough, the corresponding interpretation will score better on the operative measure than the unrestricted evolutionary tracking interpretation. Plausibly, some such measure would be best on K-max interpretationism. If so, then on K-max interpretationism experiences will in fact be distributed in accordance with that interpretation.[25] In that case, ordinary observers would dominate skeptically situated ones, possessing not only ordinary knowledge but also knowledge of astronomical data that supports theories that predict Boltzmannian domination. No cognitive instability would need ensue, as K-max interpretationism would offer ordinary observers a coherent picture of how they might have that data in a way that does not warrant distrust of it. Given that conscious observers would only arise from evolutionary processes in very large bubbles, theories that predict Boltzmannian domination would give us no reason to expect our experiences to be more orderly than we find them; hence, the order we find in our experiences would not tell against these theories.

---

[25]There are a couple of ways that K-max interpretationism could be revised in the event that there is more than one interpretation that scores best. First, if the interpretations agree, K-max interpretationism could be modified to claim that experiences are distributed in accordance with the best interpretation *or interpretations* on the K-max measure. A problem would remain if conflicting interpretations are tied for best. The second modification solves that problem. On it, *one of the best interpretations is selected at random and experiences are distributed in accordance with it*—a corollary is that if there is just one interpretation, it will be selected.

6. Going Global and Staring Down Incredulous Stares

We have just seen how, unlike some other forms of phenomenal externalism, K-max interpretationism can shield physical theories from creature-based Boltzmannian objections. One reason other theories failed was that they did not zombify large enough bubbles. As a result, they remained threatened by hordes of unfortunate observers inhabiting Boltzmann bubbles. Part of what enabled K-max interpretationism to avoid this was its embrace of *phenomenal globalism*, the thesis that subjects' experiences are directly sensitive to the entire physical state of the universe rather than fixed by some sub-total physical state of the universe. I have not argued that K-max interpretationism is the only form of phenomenal globalism that can shield physical theories from creature-based objections. I would be surprised if that were true, as the space of phenomenal globalist theories is almost wholly unexplored. Because I believe that phenomenal globalist shielding strategies merit further exploration, I will here address what I anticipate to be a common objection to phenomenal globalism: the incredulous stare.

In reply, I suggest we distinguish two thoughts that might lie behind the stare. The first is that there is a basic intuition which entails that experiences do not depend on global physical states, or at least not within universes that are as large as ours. This thought is misguided in much the way that the thought that we can establish phenomenal internalism is misguided: there is a dearth of a priori connections between physical states and experience. We can conceive of experience being fixed by physical states that are internal, external but not global, or global. Which in fact fixes them is not something we can settle by intuition.

A different thought that might lie behind the stare is that phenomenal globalism requires an a priori bizarre form of action at a distance (or else the even more a priori bizarre view that each experience occupies every region of spacetime). I admit that there is something bizarre about phenomenal globalism. However, I think that any such defect would need to be explicable in terms of some sort of more basic theoretical vice. And, on reflection, I do not see what such a vice might be. Admittedly, if a local mechanistic account of experience could be had, it might be more intelligible than any form of phenomenal globalism. But we lack such an account and the prospects for such an account seem dim. While Newton recoiled at unmediated action at a distance, quantum mechanics has taught us to accept it with natural piety. More generally, contemporary physical theories postulate physical laws that operate on the entire state of the universe at a time without presuming that their operation can be explained in terms of any more basic local mechanisms. A defense of the noted thought would need to say why what is allowed for physical laws should be forbidden for psychophysical principles.[26]

7. Global Shields and Fine-tuning

I turn now to consider some connections between the phenomenal globalist shielding strategy and fine-tuning. In particular, I'll show how using phenomenal globalism to shield physical theories from Boltzmannian objections would incur a kind of psychophysical fine-tuning, one akin to more familiar forms of (cosmological) fine-tuning. As we'll see, these forms of fine-tuning interact.

I start with some background on *cosmological fine-tuning*, the apparent fact that sets of basic physical parameters (physical constants, initial conditions, the form of the basic physical laws) take values that fall within narrow regions of parameter space that support (intelligent) life. While there is disagreement about which sets of parameters are fine-tuned, cosmological fine-tuning is widely taken to be a striking fact. I'll assume that it is such a fact, i.e. that it cries out for explanation. There is much debate in philosophy and science about how to respond to these cries. Responses include:

- Accept cosmological fine-tuning as a brute fact.
- Posit new physics that somehow derives the fine-tuned parameters as robust results of more basic parameters that are not strikingly fine-tuned.
- Posit a designer that was biased toward creating a universe with such fine-tuned parameters.

[26] See Chalmers (2010: 126). Schaffer (forthcoming) contends that the global character of fundamental laws and the local ('regional') character of the physical correlates of consciousness favors his ground-theoretic physicalist view of consciousness over naturalistic forms of dualism on which fundamental psychophysical laws generate experiences. This move is too quick: as K-max interpretationism illustrates, *local* physical differences could be systematically correlated with phenomenal differences as a consequence of a *global* psychophysical law.

- Posit a meta-law that sets first-order laws (or other parameters) and favors ones that support life.
- Posit a multiverse containing many universes that vary in parameter values, rendering it to be expected that there will be some universe with life-supporting parameters.[27]

Cosmological fine-tuning is striking because it places a special outcome (the obtaining of life-supporting parameters, in this case) within a wide class of unremarkable possible outcomes (the countless parameter settings that would not have supported life) and provides no explanation of why a special outcome obtains rather than an unremarkable one. The phenomenal globalist shielding strategy yields a form of *psychophysical fine-tuning* that shares this feature: it places a special outcome (the obtaining of a psychophysical principle that suppresses the production of Boltzmann observers and thereby shield physical theories from Boltzmannian objections) within a wide class of unremarkable possible outcomes (those involving psychophysical principles that do not shield physical theories from Boltzmannian objections).

Given that cosmological fine-tuning cries out for explanation, so too does this form of psychophysical fine-tuning. That's not to say that these two sorts of fine-tuning cry out for explanation to the same degree: it may be that, while each involve outcomes that enjoy similar degrees of intrinsic specialness, cosmological fine-tuning is more striking because it renders its special outcome far rarer among the class of possible outcomes than does the sort of psychophysical fine-tuning under consideration.[28] In any event, leaving such psychological fine-tuning unexplained would be objectionable in much the way that leaving cosmological fine-tuning unexplained would be

[27] E.g., see my (forthcoming).

[28] This is motivated by the involvement of many fine-tuned parameters in the cosmological case along with the miniscule fraction of parameter space that is estimated to be life-supporting. But the matter is not entirely clear-cut. For one, it is unclear how to estimate what proportion of the space of psychophysical principles is occupied by principles with the wanted shielding property. For another, estimates of life-supporting parameter ranges often do not take into account the possibility of Boltzmannian life, i.e. life that results from fluctuations rather than the operation of natural selection over many generations. Taking into account the possibility of Boltzmannian life would presumably widen the life-supporting ranges of parameters and, in turn, the life-supporting regions of parameter space. That result might be reversed by reconstruing cosmological fine-tuning for life as cosmological fine-tuning for *ordinary* (i.e. non-Boltzmannian) life. But this would arguably lessen the intrinsic specialness of what cosmological fine-tuning is fine-tuning *for*, and so lessen the extent to which it calls for explanation.

objectionable. Thus, proponents of the phenomenal globalist shielding strategy face a challenge: that of explaining why we're in a world that is psychophysically fine-tuned to shield physical theories from Boltzmann brains.

8. Fine-Tuning Is a Common Side Effect of Shielding

In the previous section we saw that psychophysical fine-tuning poses an explanatory challenge for the phenomenal globalist shielding strategy. If the phenomenal globalist shielding strategy were the only one that generated a fine-tuning problem, that would be a reason to reject it and opt for a different shielding strategy. However, for two reasons, the phenomenal globalist shielding strategy is not alone in generating a fine-tuning problem. First, the psychophysical fine-tuning problem for the phenomenal globalist shielding strategy is a special case of a much more general psychophysical fine-tuning problem: why is there a (partial) coincidence between the psychophysical principles that obtain and those would suppress the production of observers in skeptical or cognitively unstable situations? This problem arises for virtually any view about the realization of experience. Since it is not clear that the general psychophysical fine-tuning problem can be solved without appealing to resources that would also explain psychophysical fine-tuning for the suppression of Boltzmann observers, it is not clear that the latter provides much of a reason to reject the phenomenal globalist shielding strategy.[29]

Second, it turns out that a range of notable physics-based shielding strategies face fine-tuning problems of their own. I'll illustrate this by looking at three cases in which adjusting physical theories to ward off Boltzmannian objections results in fine-tuning and vice versa.

[29] A related point is that there also other sorts of psychophysical fine-tuning (e.g. involving a hedonic match between experiences' valence and their associated behaviors) that arise for a wide range of views—see Cutter & Crummett (forthcoming), Chalmers (2010: 132; 2020), Goff (2018), James (1890), Mørch (2018), Pautz (2006; 2020), Ross (2017), and Saad (2019*b*). It's not clear that these psychophysical fine-tuning problems can be solved without appealing to resources that would also explain psychophysical fine-tuning for the suppression of Boltzmann observers. So, it's also not clear that the latter provides much of a reason to reject the phenomenal globalist shielding strategy.

8.1 Dead Universes vs. Lively Fluctuations

Multiverse theories posit a vast ensemble of universes with varied physical parameters and initial conditions. The hypothesis that we live in *some* such multiverse is sometimes offered as an explanation of cosmological fine-tuning.[30] According to the proposed explanations, it is to be expected that there is some universe with life-supporting parameters, given a suitably large and varied multiverse. Speculative physical theories (such as eternal inflation, cosmological natural selection for blackhole production, and landscape string theories)[31] propose mechanisms for the generation of such multiverses.

An obvious way to avoid the Boltzmannian objections raised by physical theories that posit a multiverse is to reject those theories in favor of single-universe physical theories. But the latter theories typically lack an explanation of cosmological fine-tuning. So, this physics-based shielding strategy generates a fine-tuning problem.

8.2 Explaining the Thermodynamic Arrow

Our universe is subject to the second law of thermodynamics: entropy in closed systems generally does not decrease. Our universe manifests this law through entropy increase. This cries out for explanation, as it is a striking temporal asymmetry that is underpinned by (more or less)[32] temporally symmetric dynamical physical laws.[33]

One popular strategy for explaining the thermodynamic arrow invokes the *Past Hypothesis* that it's a basic fact—or basic law that—our universe started in a very low entropy macroscopic state.[34] An unpopular rival is the *Fluctuation Hypothesis* that our universe's entropic character is the result of a fluctuation away from thermal

[30] See Friederich (2017) for an overview.

[31] Respectively see Guth (2000), Smolin (1998), and Susskind (2005).

[32] In quantum field theory, parity violations and the CPT theorem jointly imply violations of time-reversal invariance. However, it's widely agreed that such temporal asymmetries are not suitable to explain the thermodynamic asymmetry (North, 2011: §3).

[33] Other aspects of the thermodynamic arrow—for instance, that it holds for subsystems of the universe as well as the universe as a whole and that it aligns with and underwrites other arrows of time—arguably stand in need of explanation as well. Exactly what else needs to be explained and how they would need to be supplemented to explain them are delicate matters that I won't explore here. But see Albert (2000) and Earman (2006) for relevant discussion.

[34] See Albert (2000) and Loewer (2012).

equilibrium.[35] In particular, the Fluctuation Hypothesis holds that a fluctuation from thermal equilibrium produced a low-entropy early state of our observable universe.

While unpopular, the Fluctuation Hypothesis has some attractive features. In some classical settings in which thermodynamics is often considered, our universe is ergodic in that it is bound to fluctuate into (or, at any rate, arbitrarily close to) a given state an infinite number of times, and hence to the posited low-entropy past state of our universe. It is an open question whether our universe is ergodic in that sense.[36] Moreover, explanations of the thermodynamic arrow that invoke the Fluctuation Hypothesis require fewer fundamental posits than those that instead invoke the Past Hypothesis: the two sorts of explanation require the same basic commitments concerning fundamental dynamical laws (and auxiliary assumptions about the distribution of microstates), but the latter sort of explanation incurs an additional basic commitment by way of the Past Hypothesis. Further, the Past Hypothesis is arguably fine-tuned: on one often cited estimate, macrostates of the sort specified by the Past Hypothesis (i.e. those that are compatible with the evolution of a universe remotely like ours) occupy only $1/10^{(10^{123})}$ of the volume of the relevant phase space.[37] If we apportion the probability accordingly, the Past Hypothesis will receive a miniscule probability, while the possibility of our universe starting in an ordinary-life-prohibitive state of thermal equilibrium will receive a very high probability (~1). And it will then appear to be very lucky that the Past Hypothesis obtained rather than that rival possibility.

The main drawback of the Fluctuation Hypothesis is that it faces Boltzmannian objections: for a given state, small fluctuations that produce it will be more prevalent than larger ones, meaning that Boltzmann brains and their ilk should be expected to dominate their ordinary counterparts. As a result, we should expect any given world that is apt to produce our universe via fluctuation to produce a preponderance of

[35] A proposal along these lines was advocated by Boltzmann (1895: 1897), which he attributed to an assistant, Ignaz Robert Schütz (1895). More recently, Chen (2021) has argued that a version of the Fluctuation Hypothesis can be defended using a move (which appeals to self-locating probabilities) required to save the Past Hypothesis's explanation of the thermodynamic arrow from the threat of future Boltzmann brains.
[36] See Chen (2021: §2.2).
[37] See Penrose (1989: 260; 2004: 730).

Boltzmann creatures via smaller and therefore more prevalent sorts of fluctuation. Thus, there is a choice between physical theories that:

> (1) explain the thermodynamic arrow by invoking the Past Hypothesis, have a fine-tuning problem, and avoid Boltzmannian objections.
> (2) explain the thermodynamic arrow by invoking the Fluctuation Hypothesis, avoid the fine-tuning problem, and face Boltzmannian objections.

It might be suggested that the Past Hypothesis is an instance of fine-tuning that can be explained by positing a multiverse. After all, a key worry about multiverse explanations of fine-tuning is that they are susceptible to Boltzmannian objections, and we've seen that the Past Hypothesis is supposed to disarm those objections on the way to explaining the thermodynamic arrow.

However, this suggestion seems untenable. A multiverse explanation of the Past Hypothesis would presumably require an ensemble that contains many universes with our laws but with entropically differing initial conditions. In particular, to render it to be expected that the Past Hypothesis holds of some universe with our fundamental dynamical laws, the ensemble would need to contain a large number and dominant proportion of such universes with high entropy initial conditions (ones at or near equilibrium). Such high entropy universes would provide more than enough opportunities for Boltzmann creatures to fluctuate into existence and dominate such a multiverse. Thus, combining the Past Hypothesis with such a multiverse explanation would render the Past Hypothesis susceptible to the very Boltzmannian objections it is supposed to avoid. On reflection, this should be unsurprising, as the posited multiverse contains an ensemble of entropically differing initial conditions that is akin to the ensemble of fluctuations that simultaneously enabled the Fluctuation Hypothesis to explain the low-entropy state of our universe and brought that hypothesis within the scope of Boltzmannian objections.

8.3 A Physics-Based Strategy for Ridding ΛCDM of Boltzmann Observers

The standard best-fit cosmological model of our universe—ΛCDM ('CDM' for cold dark matter)—holds that our universe is undergoing accelerated expansion from a low-entropy Big Bang into a flat (De Sitter) space featuring a horizon.

ΛCDM does not explain the low-entropy state of the Big Bang. It thus invites a recapitulation of the dialectic surrounding the thermodynamic arrow that we encountered above: the low-entropy state could be explained by way of the Past Hypothesis; but this posit would involve a suspect form of fine-tuning. We might rid ourselves of such fine-tuning by opting for the Fluctuation Hypothesis or by positing a multiverse with suitable variations to render unsurprising the existence of a universe like ours in which the Past Hypothesis holds. But Boltzmannian objections await those who take either of these escapes from fine-tuning. Another proposed explanation can be found in eternal inflation, which aims to explain the low-entropy state of our universe through an inflationary process that, once begun, continues without end. The inflationary process leads to a multiverse of "pocket universes" in which Boltzmann creatures dominate (on at least some cosmological measures).[38] Eternal inflation is beset with outstanding technical problems[39] and arguably itself requires fine-tuning.[40] I will set it aside and focus on a way in which ΛCDM *on its own* poses a choice between fine-tuning and the proliferation of Boltzmann creatures, i.e. *independently of the Past Hypothesis and eternal inflation*.

ΛCDM permits different coordinatizations of spacetime. Given a coordinatization, quantum mechanics issues a description of the horizon of space. On conventional coordinatizations, quantum mechanics describes the horizon in ΛCDM as involving dynamic fluctuations of the sort that would produce Boltzmann creatures.[41] However, there are also non-conventional (i.e. not-standardly-used-in-cosmology) coordinatizations. Some such coordinatizations yield descriptions on which the horizon is in a static state—it doesn't evolve with time and so doesn't involve dynamic

[38] See Guth (2000).
[39] Notable among these is the cosmological measure problem of finding a suitable measure of observers within multiverses containing infinitely many observers. An important but not easily satisfied constraint on suitable measures is that they do not result in Boltzmannian domination. This problem deserves extended consideration in connection with this paper's topic, but examining it in more detail would take us too far astray. For discussion, see Bostrom (2002*b*), Dorr & Arntzenius (2017), Smeenk (2014), and Tegmark (2014).
[40] See Carroll & Tam (2010) and Penrose (1989: 263, *passim*).
[41] See, e.g., Carroll (2020).

fluctuations—and therefore devoid of Boltzmann observers, provided that Boltzmann observers would only result from dynamic fluctuations.[42]

Thus, one can shield ΛCDM from Boltzmannian objections (insofar as they concern Boltzmann observers on the horizon) by insisting that reality conforms to the static description devoid of Boltzmann observers rather than any of the many dynamic descriptions on which they would dominate. This shielding maneuver saddles its proponents with a cosmological fine-tuning problem: why does reality conform to the (Boltzmann observer)-inhospitable description rather than the (Boltzmann observer)-friendly description? To answer this question, one could posit a multiverse with universes that vary with respect to which of these descriptions they satisfy. However, positing such a multiverse would reinstate the Boltzmannian objections: while it might rid some universes like ours of Boltzmann creatures, it would grant Boltzmann creatures reign in the multiverse at large. That would be enough to get the Boltzmannian objections off the ground, since they are no less plausible on the assumption that a multiverse is dominated by Boltzmann creatures than they are on the assumption that our universe is.

9. Fruits of the Phenomenal Globalist Shielding Strategy

We've seen that the phenomenal globalist shielding strategy has a fine-tuning problem (§7). I've argued that this objection is not decisive partly by identifying several notable physics-based shielding strategies that are themselves susceptible to fine-tuning problems (§8). Having examined these strategies, we are now in a position to appreciate several corresponding theoretical fruits that the phenomenal globalist shielding strategy can bear.

First, recall that some physical theories aim to explain cosmological fine-tuning by positing a multiverse, but in so doing they proliferate Boltzmann creatures and so face Boltzmannian objections. Unless those objections are met, they render unsatisfactory the proposed multiverse explanation of cosmological fine-tuning. Those

[42] See Boddy et al. (2016); for related proposals that appeal to Bohmian mechanics, see Goldstein et al. (2015) and Tumulka (2016).

objections can be met by combining the multiverse explanation with the phenomenal globalist shielding strategy. Provided that the multiverse is construed as an ensemble of universes that vary in their *physical* parameters and the operative global psychophysical principle is construed as applying to the multiverse itself (not just to our universe), Boltzmann creatures will be zombified throughout the multiverse and the Boltzmannian objections will collapse.[43] Thus, the phenomenal globalist strategy enables a physics-based explanation of cosmological fine-tuning—at least given that such an explanation would be in the offing if not for Boltzmannian objections.

Next, we've seen that there are several options for explaining the thermodynamic arrow. One could invoke the Past Hypothesis and maintain that there is a single universe. But this arguably requires a form of fine-tuning, given the specialness and (arguable) improbability of the initial state posited by the Past Hypothesis. Alternatively, one could invoke the Past Hypothesis and countenance a multiverse that renders the state posited by the Past Hypothesis to be expected. Or, one could appeal to the Fluctuation Hypothesis that renders an early state of our observable universe to be expected. But these options are subject to Boltzmannian objections. The phenomenal globalist strategy answers those objections on both the multiverse Past Hypothesis explanation and on the Fluctuation Hypothesis explanation. In both cases, the globalist response is that the objections fail because they require a preponderance of Boltzmann observers and the operative globalist principle zombifies the Boltzmann brains and creatures generated by the proposed explanation. As for the explanation that invokes the single-universe Past Hypothesis, the phenomenal globalist strategy neither mitigates nor exacerbates its fine-tuning problem. Thus, the phenomenal globalist strategy brightens the prospects of a physics-based explanation of the thermodynamic arrow by removing obstacles to two proposed explanations without imposing any upon their main rival.

[43] A cost of this explanation may be that it requires a kind of disuniformity in its treatment of basic parameters: whereas it construes physical parameters—or, at any rate, those that are fine-tuned in our universe—as varying across the multiverse, it construes a psychophysical parameter as invariant. Plausibly, granting that this is a cost, it is a small one that may well be worth paying if it preserves a physics-based explanation of cosmological fine-tuning.

Finally, ΛCDM accounts for an impressive range of independent astronomical observations.[44] But ΛCDM may generate a preponderance of Boltzmann creatures. Whether it does depends on which of two available sorts of quantum description reality conforms to. One can shield ΛCDM from the Boltzmannian threat by maintaining that reality conforms to the description that does not engender Boltzmann creatures. But this response is open to the charge that it requires reality to be fine-tuned so as to conform to the quantum description devoid of Boltzmann creatures rather than one on which they abound. The phenomenal globalist shielding strategy offers an alternative: by zombifying Boltzmann creatures, if such there be, it allows for ΛCDM to be construed in a way that is neutral between the two sorts of quantum descriptions while also being immune to the Boltzmannian objections. Thus, the phenomenal globalist shielding strategy enables ΛCDM to account for a range of astronomical observations without succumbing to Boltzmannian objections or requiring cosmological fine-tuning to suppress Boltzmann creatures.

To sum up, the phenomenal globalist shielding strategy incurs an initial cost in the form of fine-tuning. But this initial cost is not unique to the phenomenal globalist shielding strategy: non-globalist views about the realization of experience also seem to require some degree of psychophysical fine-tuning, and physics-based shielding strategies also suffer costs from cosmological fine-tuning. Moreover, the phenomenal globalist shielding strategy yields explanatory dividends: it brightens the prospects both for a physics-based multiverse explanation of cosmological fine-tuning and for a physics-based (Past Hypothesis or Fluctuation Hypothesis) explanation of the thermodynamic arrow; and it provides a way to embrace ΛCDM while escaping a dilemma it faces between the Boltzmannian objections and a sort of cosmological fine-tuning. Insofar as explanatory unification is a theoretical virtue, it is a point in favor of the phenomenal globalist shielding strategy that it yields the just described range of theoretical fruits in a single theoretical stroke.

[44] See https://lambda.gsfc.nasa.gov/education/graphic_history/observations.cfm for an overview.

10. Prospects for a Unified Solution

We've seen that the phenomenal globalist shielding strategy yields theoretical fruits, but that it is also subject to a psychophysical fine-tuning problem, namely that of explaining why the distribution of experience is dictated by a psychophysical principle that suppresses Boltzmann observers. A solution to that problem that coheres with the strategy would enhance the strategy's appeal. Still better would be a solution that also accounts for cosmological fine-tuning. This section explores the prospects for such a solution. I'll proceed by looking at solutions to the cosmological fine-tuning problem (from §7) and considering whether they can be extended to yield a solution of the desired sort.

Recall that one response to cosmological fine-tuning maintains that it is a brute fact. To extend this response to psychophysical fine-tuning, one need only claim that it is a brute fact that the distribution of experience is dictated by a psychophysical principle that suppresses Boltzmann observers. However, this brutalist response is dissatisfying, as it leaves striking facts unexplained. So I will set it aside.

Another approach to cosmological fine-tuning is to posit new physics that somehow renders cosmological fine-tuning to be expected as a consequence of physical parameters that are not themselves fine-tuned. However, since physical theories are not in the business of explaining psychophysical principles, new physics is not apt to account for psychophysical fine-tuning. So I will set this approach aside as well. That leaves several responses to cosmological fine-tuning: multiverse explanations, design explanations, and meta-law explanations. I'll consider each in turn.

Let's grant that a multiverse could explain cosmological fine-tuning: a multiverse containing universes with varying physical parameters could account for the existence of a universe with physical parameters that are fine-tuned for life. In that case, so too could a multiverse containing universes with varying *psycho*physical parameters explain why there is some universe with a psychophysical principle that suppresses Boltzmann observers. For given a large enough multiverse with suitable variation, some such universe would be ensured, so long as even a tiny percentage of psychophysical principles suppress Boltzmann observer production.

While the multiverse explanation of psychophysical fine-tuning may explain why some universe is psychophysically fine-tuned to suppress Boltzmann observers, that explanation does not cohere with the phenomenal globalist shielding strategy. To see this, note that the multiverse explanation of cosmological fine-tuning faces a challenge: why do we observe a universe that is cosmologically fine-tuned, given that most universes are not cosmologically fine-tuned? The going response to this challenge appeals to an observation selection effect. The response is that in a multiverse of the posited sort, it's to be expected that intelligent observers will find themselves in such a fine-tuned universe—after all, it is precisely those universes that are hospitable to such observers. Hence, it's to be expected that we will find ourselves in such a universe if we are in such a multiverse. A parallel challenge applies to the multiverse explanation of psychophysical fine-tuning: why do we observe a universe that is psychophysically fine-tuned, given that most universes are not? Unlike the original challenge, this challenge cannot be met by appealing to an observation selection effect. The trouble is that in a multiverse with the posited psychophysical variation observers are not confined to psychophysically fine-tuned universes—indeed, we would expect the preponderance of them to be in universes governed by principles that do not suppress the production of Boltzmann observers.[45]

Next, there are meta-law explanations of cosmological fine-tuning. These explanations typically appeal to an *axiarchic* meta-law. The axiarchic meta-law constrains the setting of cosmological parameters so as to promote a certain (normative) value. The explanation maintains that that value is best (or at least sufficiently well) realized through the existence of intelligent life. Thus, given the axiarchic meta-law and cosmological parameters that only permit intelligent life within narrow ranges, it's to be expected that those parameters will in fact take the values in those ranges and so result

[45] One could ensure that observers in the multiverse generally find themselves in a universe psychophysically fine-tuned for Boltzmann observer suppression by maintaining that the psychophysical variation across the multiverse is between different principles that suppress such observers. While this would answer the challenge, it would also result in a multiverse that is itself psychophysically fine-tuned to suppress Boltzmann observers—meaning that the multiverse explanation would fail in its ambition to explain fine-tuning with a posit that is not fine-tuned.

in a universe that's cosmologically fine-tuned. A slightly more general axiarchic meta-law of this sort would apply not only to cosmological parameters but also to psychophysical parameters, biasing both toward the realization of a certain value. It would not be especially surprising if a value whose realization requires intelligent life also favors scenarios where those observers are not skeptically situated. These considerations suggest what we might call:

> *The unified axiarchic hypothesis*: A meta-law constrains the cosmological and psychophysical parameters in a manner that is biased toward the promotion of a certain value. That value favors both the existence of intelligent observers and their having orderly experiences in non-skeptical scenarios. These facts explain the cosmological fine-tuning of physical parameters for intelligent observers, the prevalence of orderly experiences among observers, and why Boltzmann brains and creatures are (or would be) unconscious.

In contrast to the multiverse explanation of psychophysical fine-tuning, the unified axiarchic hypothesis fits with the phenomenal globalist shielding strategy. Insofar as phenomenal globalist principles are otherwise natural candidates for psychophysical principles that would suppress Boltzmann observers, they are the sort of psychophysical principles that would be explained by the unified axiarchic hypothesis. And because the unified axiarchic hypothesis does not predict that there are universes with other sorts of psychophysical principles as well, its explanation does not generate a puzzle as to why we find ourselves in a universe that is psychophysically fine-tuned.

Finally, there are design explanations of cosmological fine-tuning. These explanations appeal to a designer who constrains physical parameters in accordance with some value that favors the existence of intelligent observers. If such an explanation works, it can be straightforwardly extended to explain psychophysical fine-tuning by supposing that the designer also sets the psychophysical parameters in a manner that is biased against the production of Boltzmann observers. This extension is a natural one, as it is plausible that values that favor the existence of intelligent observers would also favor the existence of intelligent observers whose experiences are orderly and whose circumstances are non-skeptical.

Extending the design and meta-law explanations in this fashion increases the range of data they account for without significantly augmenting the material they use to explain them. Insofar as explanatory power is a virtue, this is a point in favor of the extended explanations. Of course, like the explanations from which they are extended, they will remain highly controversial. Those who antecedently embrace such explanations may extend them at little or no additional cost to account for psychophysical fine-tuning for Boltzmann observer suppression and thereby gain access to the fruits of the phenomenal globalist shielding strategy.[46] For those who are at least open to these explanations, they offer a proof of concept for how the psychophysical fine-tuning generated by the phenomenal globalist shielding strategy might be explained. Those who reject design and meta-law explanations of psychophysical fine-tuning are under pressure to reject the corresponding explanations of cosmological fine-tuning along with the phenomenal globalist shielding strategy. This leaves such theorists with a severely limited range of responses to the challenge of simultaneously answering the Boltzmannian objections and accounting for cosmological fine-tuning.

11. Conclusion

We've seen that scientists have resorted to exotic posits in their physical theorizing in order to shield their theories from Boltzmannian objections. This paper has developed an alternative strategy that shields physical theories from Boltzmannian objections through philosophical adjustments. In particular, the strategy zombifies Boltzmann brains and Boltzmann creatures by invoking a suitable phenomenal globalist principle that zombifies such entities while leaving the consciousness of ordinary observers intact. While the proposed strategy turned out to have a psychological fine-tuning problem, it also turned out that the strategy helps with fine-tuning problems in physics and that the psychophysical fine-tuning problem admits of solutions. In light of this, I suggest that the phenomenal globalist shielding strategy deserves to be taken seriously alongside physics-based shielding strategies. Adjudicating between these strategies is a

[46] For discussion of theistic vs. naturalistic responses to Boltzmann brains in the context of thermodynamics, see Monton (2018).

task for further research. For now, I conclude by highlighting some methodological morals.

The phenomenal globalist shielding strategy uses philosophical assumptions to suppress the production of Boltzmann observers without suppressing the production of Boltzmann brains. This does not show that physics-based strategies for suppressing Boltzmann brain production should be abandoned. But it does show that avoiding the proliferation of Boltzmann brains should not be unreflectively endorsed as a criterion of adequacy for physical theories.

The phenomenal globalist shielding strategy shows how phenomenal globalism can help answer Boltzmannian objections to physical theories. Since some of these theories enjoy empirical support, I take this to motivate phenomenal globalism. Herein lie potential evidential connections that span physics, neuroscience, and philosophy of mind. Evidence for physical theories that proliferate Boltzmann brains may indirectly tell in favor of phenomenal globalism (per its ability to shield those theories) over phenomenal internalism and mild forms of phenomenal externalism (per their inability to shield those theories). In the other direction, evidence for phenomenal internalism or mild forms of externalism may tell against phenomenal globalism and, in turn, against the phenomenal globalist shielding strategy and physical theories it would shield from Boltzmannian objections.[47]

Reflecting on Boltzmannian objections has revealed a complex interplay between the fine-tuning problems in physics and their less familiar psychophysical counterparts. Some physical approaches to cosmological fine-tuning face Boltzmannian objections and so invite a (phenomenal globalist) shielding strategy that requires psychophysical fine-tuning. This reveals a way in which attempts to rid theories of fine-tuning can fail by instead relocating it. The phenomenal globalist strategy at once suffers a psychophysical fine-tuning problem and illustrates why this sort of problem is not decisive. For one, a posit that generates psychophysical fine-tuning may enable physical

[47] For example, phenomenal-neural structural correlations arguably support phenomenal internalism—see Pautz (2013) for an overview of relevant neuroscientific findings and an argument that uses them against certain forms of phenomenal externalism.

explanations of cosmological fine-tuning. For another, some explanations of cosmological fine-tuning—notably the design and meta-law explanations—can be straightforwardly and plausibly extended to explain psychophysical fine-tuning. On the other hand, some explanations of cosmological fine-tuning—notably, the multiverse explanation—resisted such extensions.

All this suggests that the problems posed by Boltzmann brains, cosmological fine-tuning, and psychophysical fine-tuning are deeply intertwined, and that a fruitful constraint on solving any of these problems may be coherence with satisfactory solutions to the other problems.

References

Albert, D. (2000). *Time and chance*. Harvard.

Barrett, J.A. (1996). Empirical adequacy and the availability of reliable records in quantum mechanics. *Philosophy of Science*, 63(1):49–64.

Boddy, K. K., & Carroll, S. M. (2013). Can the Higgs Boson Save Us From the Menace of the Boltzmann Brains?. arXiv preprint arXiv:1308.4686.

Boddy, K. K., Carroll, S. M., & Pollack, J. (2016). De Sitter space without dynamical quantum fluctuations. *Foundations of Physics*, 46(6), 702-735.

Boltzmann, L. (1895). On certain questions of the theory of gases. *Nature*, 51(1322), 413-415.

Boltzmann, L. (1897). Über die Unentbehrlichkeit der Atomistik in der Naturwissenschaft. *Annalen der Physik*.

Bostrom, N. (2002*a*). *Anthropic Bias: Observation Selection Effects} in Science and Philosophy*. Routledge.

Bostrom, N. (2002*b*). Self-Locating Belief in Big Worlds: Cosmology's Missing Link to Observation. *Journal of Philosophy* 99 (12):607-623.

Carlip, S. (2007) Transient Observers and Variable Constants, or Repelling the Invasion of the Boltzmann's Brains, *JCAP* 06 001 [hep-th/0703115] [INSPIRE].

Cappelen, H., & Dever, J. (2021). Making AI intelligible: Philosophical foundations (p. 192). OUP.

Carroll, S. M. (2020). Why Boltzmann brains are bad. In S. Dasgupta & B. Weslake (Eds.) *In Current Controversies in Philosophy of Science*. Routledge.

Carroll, S. M., & Tam, H. (2010). Unitary evolution and cosmological fine-tuning. arXiv preprint arXiv:1007.1417.

Chalmers, D. (2018). Structuralism as a response to skepticism. *The Journal of Philosophy*, 115, 625–660

Chen, E.K. (2021). Time's Arrow and Self-Locating Probability. *Philosophy and Phenomenological Research*.

Chalmers, D. (2010) *The Character of Consciousness*. OUP.

Chalmers, D. (2020). Debunking Arguments for Illusionism about Consciousness. *Journal of Consciousness Studies*, 27(5-6), 258-281.

Chalmers, D. (2022) *Reality+: Virtual Worlds and the Problems of Philosophy*. Norton.

Crawford, L. (2013). Freak observers and the simulation argument. *Ratio*, 26(3), 250-264.

Cutter, B. & D. Crummett (forthcoming). Psychophysical Harmony: A New Argument for Theism. In *Oxford Studies in Philosophy of Religion*.

Dainton, B. (2012). On singularities and simulations. *Journal of Consciousness Studies*, 19(1-2), 42-85.

Dalbey, B. & B. Saad. (2022) "Internal constraints for phenomenal externalists: a structure matching theory." *Synthese* 200.5:1-29.

Donald Davidson (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association* 60 (3):441-458.

De Simone, A., Guth, A. H., Linde, A., Noorbala, M., Salem, M. P., & Vilenkin, A. (2010). Boltzmann brains and the scale-factor cutoff measure of the multiverse. *Physical Review* D, 82(6), 063520.

Dogramaci, S. (2020). Does my total evidence support that I'm a Boltzmann Brain?. *Philosophical Studies*, 177(12), 3717-3723.

Dorr, C., & Arntzenius, F. (2017). Self-locating priors and cosmological measures. *The philosophy of cosmology*, 396-428.

Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.

Earman, J. (2006). The "past hypothesis": Not even false. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 37(3), 399-430.

Elga, A. (2004). Defeating Dr. Evil with self-locating belief. *Philosophy and Phenomenological Research*, 69(2)383-396.

Friederich, S., "Fine-Tuning", *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), E.N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/fine-tuning/>.

Gendler, T. S., & Hawthorne, J. (2005). The real guide to fake barns: A catalogue of gifts for your epistemic enemies. *Philosophical Studies*, 331-352.

Goff, P. (2018) Conscious thought and the cognitive fine-tuning problem, *Philosophical Quarterly*, 68 (270), pp. 98–122
Goldstein, S., Struyve, W., & Tumulka, R. (2015). The Bohmian approach to the problems of cosmological quantum fluctuations. arXiv preprint arXiv:1508.01017.
Guth, A.H. (2000). Inflation and eternal inflation. *Physics Reports*, 333, 555-574.
James, W. (1890) *The Principles of Psychology, Vol. 2*. New York: Henry Holt and Company
Kotzen, M. (2020). What Follows from the Possibility of Boltzmann Brains?. In S. Dasgupta & B. Weslake (Eds.) *Current Controversies in Philosophy of Science.* (pp. 21-34). Routledge.
Lewis, D. (1974). Radical interpretation. *Synthese* 27 (July-August):331-344.
Lewis, D. (1994). Reduction of mind. In S. Guttenplan (ed.), *Companion to the Philosophy of Mind*. Blackwell.
Loewer, B. (2012) "The Emergence of Time's Arrows and Special Science Laws from Physics", *Interface Focus*, 2(1): 13–19.
Mørch, H.H. (2018) The evolutionary argument for phenomenal powers, *Philosophical Perspectives*, 31, pp. 293–316.
Monton, B. (2018). Atheistic Induction by Boltzmann Brains. In J. Walls & T. Dougherty (eds.), *Two Dozen (or so) Arguments for God: The Plantinga Project*. OUP.
Myrvold, Wayne C. (2016). Probabilities in statistical mechanics. In A. Hajek & C. Hitchcock (Eds.), *The Oxford Handbook of Probability and Philosophy* (pp. 573–600). OUP.
Norton, J.D. (2020). You are not a Boltzmann Brain. Manuscript. URL: http://philsci-archive.pitt.edu/id/eprint/17689/contents
North, J. (2011). Time in thermodynamics. In *The oxford handbook of philosophy of time.*
Nozick, R. (1981). *Philosophical Explanations*. Harvard.
Page, D.N. (2006). The Lifetime of the Universe. *Journal of the Korean Physical Society*, 49(9), 711.
Page, D.N. (2008). Is our universe likely to decay within 20 billion years?. *Physical Review* D, 78(6), 063535.
Pautz, A. (2006) Sensory awareness is not a wide physical relation: An empirical argument against externalist intentionalism, *Noûs*, 40 (2), pp. 205–240.
Pautz, A. (2013). The real trouble for phenomenal externalists: New empirical evidence for a brain-based theory of consciousness. In R. Brown (Ed.), Consciousness inside and out: Phenomenology, neuroscience, and the nature of experience (pp. 237–298). Berlin: Springer
Pautz, A. (2014). The Real Trouble for Armchair Arguments Against Phenomenal Externalism. In M. Sprevak & J. Kallestrup (eds.), *New Waves in Philosophy of Mind*. Palgrave Macmillan. pp. 153-181.
Pautz, A. (2020). Consciousness and Coincidence: Comments on Chalmers. *Journal of Consciousness Studies* (5-6):143-155
Pautz, A. (2021). Consciousness meets Lewisian interpretation theory: A multistage account of intentionality. In Uriah Kriegel (ed.), *Oxford Studies in Philosophy of Mind*, Vol. 1.
Penrose, R. (1989). Difficulties with Inflationary Cosmology. *Annals of the New York Academy of Sciences*, 571(1), 249-264.
Penrose, R., (2004) *The Road to Reality: A Complete Guide to the Laws of the Universe*, London: Vintage.
Ross, J. (2017). Idealism and Fine-Tuning. T. Goldschmidt & K.L. Pearce (eds), *Idealism: New Essays in Metaphysics*. OUP.
Saad, B. (2019*a*). Spatial experience, spatial reality, and two paths to primitivism. *Synthese*, 1-23.
Saad, B. (2019*b*). A teleological strategy for solving the meta-problem of consciousness. *Journal of Consciousness Studies*, 26(9-10), 205-216.
Saad, B. (2022) Cognitive Phenomenology Meets Dualism. Manuscript.
Saad, B. (forthcoming) Fine-Tuning Should Make Us More Confident that Other Universes Exist. *American Philosophical Quarterly.*
Smeenk, C. (2014). Predictability crisis in early universe cosmology. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 46 (1):122-133.
Smolin, L. (1998). *The life of the cosmos*. OUP.
Susskind, L. (2005). *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design.* Brown and Company: Little
Tegmark, M. (2014). *Our Mathematical Universe*. New York: Knopf.
Tumulka, R. (2016). Long-time asymptotics of a Bohmian scalar quantum field in de Sitter space-time. *General Relativity and Gravitation*, 48(1), 1-10.

Tye, M. (1995). *Ten Problems of Consciousness*. MIT Press.
Weatherson, B. (2003). What good are counterexamples?. *Philosophical Studies*, 115(1), 1-31.
Williamson, T. (2000). *Knowledge and its Limits*. OUP.
Williamson, T. (2007). *The Philosophy of Philosophy*. Wiley-Blackwell.
White, R. (2000). Fine-tuning and multiple universes. *Noûs*, 34(2), 260-276.

Funding: This work was supported by the Sentience Institute, Utrecht University, and the John Templeton Foundation. The views expressed are the author's and do not necessarily reflect those of funders.