Boltzmann Brains

Isaac Wilhelm

Penultimate draft of Jan 3, 2026.

Please cite the version in S. O. Hansson & A. Wilson (Eds.), *Comprehensive Philosophy of Science*

Abstract

This paper reviews philosophical and physical theories surrounding Boltzmann Brains. As numerous calculations, proofs, and plausibility arguments suggest, with extremely high probability, there might exist many disembodied beings—namely, Boltzmann Brains—with experiences subjectively indistinguishable from yours. In fact, the vast majority of intelligent observers in the universe are Boltzmann Brains, if certain physical theories or models are right. This generates a problem: perhaps you are far more likely to be a Boltzmann Brain than an ordinary observer. In response to this problem, philosophers and physicists have proposed new principles for rational credences, for how evidence confirms theories, for physical laws, and more. In what follows, I summarize several of those proposals, highlighting both their advantages and their shortcomings.

1 Introduction

Consider all the intelligent observers who exist throughout the universe's history. Some contemporary theories of cosmology, thermodynamics, and statistical mechanics, suggest that the vast majority of these observers popped into existence more-or-less randomly. They did not evolve into existence, in accord with the mechanisms of evolutionary biology. Nor were they engineered in the way

that, for example, sentient robots might be. Instead, these intelligent observers were generated arbitrarily, by fluctuations in the universe's physical state.

There is great variety in what the lives of these intelligent observers are like. Some fluctuate into existence and then immediately fluctuate back out, their physical constituents almost instantaneously dissipating into the void. Others fluctuate into existence and stick around for a while. Many of these longer-lived observers are brain-like structures floating alone in otherwise empty space. Nevertheless, they are stable enough to undergo the sorts of electrical and chemical fluctuations that the brains of actual human beings—like you and I, presumably—undergo, in the course of living a full-fledged life. Many of these intelligent observers have memories of growing up wherever you were raised, attending whatever schools you attended, and doing everything else that you have done. Of course, for these intelligent observers, none of that actually happened: their memories—unlike, presumably, yours—resulted from random electrical and chemical fluctuations. But their subjective experiences are indistinguishable from your own. So for all you can tell, you are one of these intelligent observers who only recently fluctuated out of the energetic emptiness of space.

This disturbing line of thought raises what is called the 'Boltzmann Brain problem'; the intelligent observers at issue are often called 'Boltzmann Brains'. Slightly more precisely, a Boltzmann Brain is a configuration of matter that, along with its local environment, is as close as possible to thermal equilibrium while still qualifying as an intelligent observer (Carroll, 2021, p. 10). These observers are named after the physicist Ludwig Boltzmann, whose theories of thermodynamics and statistical mechanics contained the physical ideas used to argue that Boltzmann Brains almost certainly exist. Those same physical ideas, it turns out, can be used to argue that Boltzmann Brains far outnumber what are often called 'ordinary observers' – intelligent beings who arose through more familiar mechanisms, such as the biological mechanisms which drive evolution, and more generally the thermodynamic mechanisms which seem to have guided the universe for about 13.8 billion years.

The Boltzmann Brain problem arises from the fact that many Boltzmann Brains have experiences subjectively indistinguishable from any given ordinary observer's experiences. Just as you are currently having the experience as of

sitting in front of a computer, say, many Boltzmann Brains have the experience as of sitting in front of a computer too. And while the experiences of ordinary observers are generally veridical – there really is a computer in front of you, presumably – the experiences of Boltzmann Brains are not – there is, in fact, no computer in front of the Boltzmann Brains having a subjectively indistinguishable experience.

That raises a question. Given your evidence, should you think that you are (i) one of the comparatively few ordinary observers, having veridical experiences to which that evidence corresponds, or (ii) one of the comparatively many Boltzmann Brains, having subjectively indistinguishable yet non-veridical experiences to which that evidence corresponds? Since the number of Boltzmann Brains satisfying (ii) is far, far greater than the number of ordinary observers satisfying (i), it seems rational to conclude that you are a Boltzmann Brain rather than an ordinary observer.

Hence the Boltzmann Brain problem. You seem to have good reason for believing the skeptical hypothesis that you are a Boltzmann Brain floating in the void rather than an ordinary observer. For your subjective state is compatible with being either a Boltzmann Brain or an ordinary observer. And at least some reasonable physical theories suggest that the former far outnumber the latter. So there is rational pressure to think that you are probably a Boltzmann Brain, and so to draw the skeptical conclusion that none of your experiences have been veridical.

In this paper, I provide an overview of different responses to the Boltzmann Brain problem. In Section 2, I present the problem in more detail: I explain how considerations based on entropy, the second law of thermodynamics, microstates, and macrostates, suggest the existence of Boltzmann Brains; then I formulate the problem precisely. In Section 3, I discuss responses to the problem which claim that theories proliferating Boltzmann Brains are self-undermining. In Section 4, I discuss responses to the problem which claim that even if there are many Boltzmann Brains, you should not think that you are one of them. In Section 5, I discuss responses which accept the problem's conclusion: perhaps you are indeed a Boltzmann Brain.

2   Setting Up the Problem

In this section, I summarize the physical ideas which suggest the existence of Boltzmann Brains.[1] In Section 2.1, I provide an explanation—designed to avoid presupposing lots of physics—of what entropy is, how entropy and energy relate, how entropy tracks macroscopic features of the world, how all this connects to the second law of thermodynamics, and finally, how all that supports physical arguments for the existence of Boltzmann Brains. In Section 2.2, I present a precise version of the Boltzmann Brains problem, and I explain how that problem differs from the simulation argument.

2.1   Physical Arguments for Boltzmann Brains

Historically, concerns about Boltzmann Brains arose from theories of statistical mechanics, and in particular, from analyses of the second law of thermodynamics. Roughly put, the second law claims that a particular quantity—called 'entropy'—increases or remains constant over time. This is a rough description for several reasons, one of which is the controversy over exactly how to interpret what the second law says.

One interpretation, stemming from work due to Boltzmann (1877/2015), takes the second law to be a statistical generalization: so the law claims merely that with high probability, entropy increases or remains constant over time. An influential argument for this interpretation employs a particular definition of entropy. But before presenting that definition, it is worth sketching a rough but intuitive gloss of what entropy is.

According to the rough but intuitive gloss, a system's entropy tracks something like the distribution of energy across that system, which in turn correlates with the extent to which energy is available for work. The energy distribution of a higher-entropy system is more uniformly distributed, to put it roughly, than

---

[1]The summary focuses, in both its examples and its general exposition, on physical ideas developed in the study of classical mechanics. For more up-to-date presentations of the physical mechanisms which may generate Boltzmann Brains, including discussion of spontaneous fluctuation in de Sitter space, see (Boddy et al., 2017; Bousso & Freivogel, 2007).

the energy distribution of a lower-entropy system. When energy is more uniformly distributed, it is less organized—that is, more evenly spread out—and so less amenable for being used: it is harder for creatures like us, working with our macroscopic tools, to convert that energy into mechanical work. When energy is less uniformly distributed, it is more organized—that is, less evenly spread out—and so more amenable for being used: it is easier for creatures like us, working with our macroscopic tools, to convert that energy into mechanical work. For when energy clumps up here and there throughout the system, those clumps of energy can sometimes be leveraged to push stuff around. When one part of a system contains significantly more energy than another part of the system, we can exploit that energy differential—across the two parts—to get macro-sized stuff moving. With a piston placed at just the right spot in a low-entropy gas, for instance, the higher-energy part of the gas can push against one side of the piston, thereby overcoming any pushing which the low-energy part of the gas might exert on the other side of the piston, and so the piston moves: and that motion can be used to turn gears, rotate levers, and so on.[2] When a piston is placed most anywhere in a high-entropy gas, however, the different parts of the gas push against the two sides of the piston to roughly the same degree—since they have roughly the same amount of energy, and more generally, the same macroscopic properties—and so the piston does not move. Hence the rough but intuitive gloss of entropy, according to which entropy tracks energy distribution, which in turn corresponds to the availability of energy for work.[3]

---

[2]This particular example, of doing work by exploiting a non-uniform energy distribution, makes additional assumptions about the macrostates of the gas's parts: for instance, that in the high-energy region, most of the particles are moving in a common direction, so that the piston can be put in their path and so the particles' energy can get the piston moving.

[3]A metaphor due to Feynman (1965, pp. 119-120) offers another way of thinking about what entropy is like. Imagine sitting on a few towels at the beach. Suddenly it starts to rain. You grab the towels and run to shelter, getting wet in the process, and start using the towels to dry off. The towels are wet, but not uniformly: the distribution of water across each towel is non-uniform, and so in that sense, more 'organized'; this bit has a lot of water, this other bit does not, and so on. You start drying off by using one of the drier parts of a towel. Once that part becomes approximately as wet as the parts near it—once the local distribution of dampness, in other words, becomes approximately uniform there—you move on to another dry patch of a towel, and use it instead. This is kind of what entropy is like. The water is like energy. The non-uniform distribution of water across the towels, when you reach the shelter, is like a low-entropy state of a system. The fact that you can use this non-uniform distribution to make yourself drier is like the fact that we can use the non-uniform distribution of energy in

Non-uniform energy distributions come equipped with all sorts of macroscopic structure. For example, suppose Susie takes a sip of coffee. There is a non-uniform energy distribution which exhibits precisely the macroscopic properties of a biological organism – namely, Susie – lifting up a cup of warm coffee – which itself consists of various macroscopic properties – and taking a drink. Or take me typing on this computer. There is another non-uniform energy distribution which exhibits precisely the macroscopic properties corresponding to a being with thus-and-so biological and thermodynamical features – namely, me – pressing buttons on an object with various electrical, chemical, and thermodynamic properties of its own – my computer. So facts about entropy track facts about energy distributions, which in turn correlate with various arrangements of macroscopic properties – and therefore of ordinary material objects, doing this or that, at the macroscopic level.

It is worth presenting a somewhat more precise characterization of entropy. Take any physical system: a gas in a box, the universe, whatever. This system has an associated state space:[4] that is, a set of microstates which that system might have, each of which represents a physically possible distribution of microphysical properties—the positions and momenta of the system's particles, say—across the system. Every microstate corresponds to a macrostate: that is, a physically possible distribution of macroscopic properties—the pressure, temperature, volume, and so on, of the system's various macro-sized parts—across the system. When a system has a particular microstate, the system has a macrostate too: namely, the macrostate corresponding to that microstate. In addition, each macrostate corresponds to a region of state space: the set of all microstates which generate the same physically possible distribution of macroscopic properties. Each region has a particular state space volume, quantified using a mathematical measure.

All this machinery can be used to define the entropy of a system. In particular, the entropy of a system with a specific microstate is a function of the volume of the state space region associated with the macrostate to which that microstate

a low-entropy system to do mechanical work. And the fact that you cannot dry off effectively using a completely damp towel is like the fact that we cannot effectively use the energy in a high-entropy system to do work.

[4]The state space could be a phase space, if the system is classical, or a Hilbert space, if the system is quantum mechanical.

corresponds.[5] In slogan form: a system's entropy, when the system has a particular microstate, is given by the volume of the state space region—corresponding to the system's macrostate—which contains that microstate.

The more precise characterization of entropy just given—which invokes microstates, macrostates, state space regions, and so on—helps explain the legitimacy of the rough but intuitive gloss of entropy given earlier. To see how, consider a gas confined to a box. Then here is a noteworthy physical fact: a macrostate in which energy is non-uniformly distributed throughout the gas corresponds to relatively few microstates, and so occupies a small volume in the gas's state space. For example, there are fewer ways to have (i) a large amount of the gas's total energy clumped in the left half of the box and a small amount of that energy clumped in the right half of the box, than there are ways to (ii) have the gas's energy distributed uniformly throughout the box as a whole. So when the gas's macrostate exhibits a non-uniform distribution of energy, the gas has low entropy. And as explained earlier, non-uniform distributions of energy can be used to do mechanical work. So a gas's entropy is low just in case the volume of the macroscopic state space region containing the gas's microstate is also low – by the more precise characterization of entropy – which holds just in case the energy distribution across the gas is non-uniform, in which case the gas's energy is available for work – which is the rough but intuitive gloss of entropy.

To further illustrate the connection between the rough but intuitive gloss and the more precise characterization, consider the region of state space which contains microstates corresponding to systems in macroscopic equilibrium. When a system's microstate is in this region, that system's energy distribution is maximally uniform. So according to the rough but intuitive gloss, that system's entropy is maximal. In addition, here is another noteworthy physical fact: the state space volume of this equilibrium macrostate region is larger, by many orders of

---

[5]More formally, let x be a microstate of the system, let $M(x)$ is the macrostate to which x corresponds, let $R_{M(x)}$ be the region of state space containing all the microstates corresponding to $M(x)$, let μ be a measure appropriate for the state space at issue, and let k_B be Boltzmann's constant. Then the entropy $S(x)$ of the system is defined by Boltzmann's equation

$$S(x) = k_B \ln\big(\mu(R_{M(x)})\big)$$

For more details, see (Stowe, 2007, pp. 126-130).

magnitude, than the state space volumes of all other regions corresponding to all other macrostates which the system can have. So according to the more precise characterization, when a system's microstate is in this region, that system's entropy is maximal. And so both the rough but intuitive gloss, and the more precise characterization, associate maximal entropy with the equilibrium macrostate.

As it turns out, the machinery of microstates, macrostates, state space regions, the more precise definition of entropy, and so on, can be used to give an extremely elegant physical account of the second law of thermodynamics. A detailed presentation of that account is beyond the scope of this paper; but very roughly put, the basic idea is this.[6] Once again, consider a gas in a box. Suppose that initially, the gas has low entropy: corresponding, say, to the gas being entirely within the box's left half. Since the gas is concentrated entirely to the left, all of the gas's energy is clumped to the left too. So the gas's energy distribution is highly non-uniform. And so it follows that the volume of the macrostate state space region containing the gas's microstate is small. Actually, that is an understatement. When the gas has low entropy, the volume of its macrostate state space region is ridiculously, almost unimaginably small compared to many other macrostate regions into which the gas's microstate may evolve: for a dilute gas with only 10^{20} particles, the volume of the region corresponding to the equilibrium macrostate is roughly $10^{10^{20}}$ times greater than the volume of the region corresponding to the macrostate of the gas being confined to the box's left half (Goldstein, 2001, p. 43). So with extremely high probability, as the gas's microstate evolves in accord with the dynamical laws, that microstate will make its way from the much smaller regions – associated with the gas having lower entropies – to much larger regions – associated with the gas having higher entropies. With high probability, the gas's energy distribution will spread out: more colloquially put, with high probability, the gas will expand to evenly fill the container. Hence the present interpretation of the second law of thermodynamics: with high probability, entropy increases or remains the same.

Note that this interpretation accounts for the macroscopic, thermodynamic behavior of systems by appealing to those systems' microstates. It accounts for

---

[6]For more thorough exposition, see (Albert, 2000; Goldstein, 2001; Lazarovici & Reichert, 2015).

the entropy-increasing evolution of the gas in a box, for instance, by appealing to microphysical facts about the particles comprising the gas, the dynamical laws governing those particles' evolution, and so on. Historically, Boltzmann explained macroscopic evolution by appealing to the statistical behavior of microphysical states evolving in accord with microscopic equations of motion. This is an incredible accomplishment: it proposes to explain, and in that sense reduce, a wide variety of macroscopic goings-on to the microphysical world.[7] The reduction appeals to claims about the microdynamics of the systems at issue, and also to probabilistic claims about how those systems' microstates are likely to evolve.

And these sorts of claims, about microdynamics and probability, form the basis of arguments for the existence of Boltzmann Brains. The theoretical apparatus which supports such attractive accounts of entropy, the second law, and other macroscopic goings-on—and which has passed experimental tests again and again and again—also supports arguments for the view that Boltzmann Brains exist. Let us see why.

For starters, note that the entropy of a universe whose macrostate features Boltzmann Brains, rather than ordinary observers, is far greater than the entropy of a universe whose macrostate features ordinary observers rather than Boltzmann Brains. For the energy distributions which Boltzmann brains require, while not maximally uniform, are still far more uniform than the energy distributions which ordinary observers require. The kind of energy clumping associated with the macroscopic structures of ordinary observers on a planet like Earth, in a galaxy like the Milky Way, is astronomically more non-uniform than the kind of energy clumping associated with the macroscopic structures of Boltzmann Brains without any planets or galaxies at all. So because of the connection between entropy, energy distribution, and state space regions outlined earlier, the regions of state space containing microstates whose corresponding macrostates feature only Boltzmann Brains have far larger volume than the regions of state space contain-

---

[7]One of the most important features of this concerns its connection to the arrow of time: the probabilistic interpretation of the second law supports an account of how the time-asymmetric character of the macroworld—the fact that people tend to get older rather than younger, that ice cubes tend to get warmer rather than get colder, that gases tend to fill their containers rather than contract to a corner, and so on—can be derived, along with some other assumptions, from time-symmetric microphysical laws (Lebowitz, 1993b, 1993a).

ing microstates whose corresponding macrostates feature ordinary observers in an ordinary galaxy.

It follows that if the universe's state space satisfies certain specific conditions, and the dynamical laws governing the universe's microstate have certain properties, the probability of the universe evolving to a microstate containing Boltzmann Brains is more-or-less one. The 'certain specific conditions' vary, depending on exactly which physical theory or model of the universe is being explored. To have a somewhat concrete example, suppose that the universe is a classical system consisting of point particles, and suppose that the universe's microstate evolves, according to the dynamical laws, within a bounded phase space (Carroll, 2021, p. 8).[8] Then various mathematical and physical arguments suggest that with extremely high probability, the universe will produce Boltzmann Brains.[9]

Here is a brief sketch of three such arguments. First, as Poincaré showed, most any initial microstate will return, when evolved forward in accord with the laws, arbitrarily close to where it began (1890). So if the universe ever passes through a phase space region corresponding to a macrostate which features Boltzmann Brains—which seems plausible, given how big those regions are—then over sufficiently long periods of time, the universe will return to that macrostate repeatedly. Second, ergodicity results—which, though difficult to establish in general, have been produced for some models—show that if certain 'mixing' conditions obtain, then with probability one, the amount of time which the universe's microstate spends in any given region of phase space will be proportional to the size of that region. Since regions featuring Boltzmann Brains are far larger than regions featuring ordinary observers, it follows that if the relevant conditions obtain, then the universe's microstate will spend far more time in (i) regions corresponding to macrostates featuring only Boltzmann Brains, than (ii) regions corresponding to macrostates featuring ordinary observers instead (Petersen, 1983; Reichert, 2024). Third, since the second law assigns a tiny but non-zero probability to entropy de-

---

[8]Analogous results hold if the universe is a quantum system instead, with the dynamical laws of quantum physics: there too, with probability around one, the universe will exhibit lots and lots of Boltzmann Brains throughout its evolution (Cotler et al., 2022; Dyson et al., 2002). Note especially that even though de Sitter space may be globally unbounded, Dyson et al. argue that each observer's causal patch—bounded by a de Sitter horizon—has finite entropy, so certain Poincaré recurrence results hold there (2002, p. 2).

[9]Though for concerns about the extent to which these arguments succeed, see (Norton, 2022).

creasing, it follows that over sufficiently long stretches of time, entropy-decreasing fluctuations are virtually guaranteed to occur repeatedly. And since the sizes of regions featuring only Boltzmann Brains are so much larger than the sizes of regions featuring only ordinary observers, numerous typicality arguments suggest that typical entropy-decreasing fluctuations will generate Boltzmann Brain macrostates far more often than ordinary observer macrostates (Albrecht, 2004; Albrecht & Sorbo, 2004; Carroll, 2010; Eddington, 1931).[10]

That is the physical basis of Boltzmann Brains. Differences in entropy correspond to differences in energy distribution, which in turn correspond to differences in macrostates. More uniform energy distributions, which have relatively few macroscopic structures, correspond to large regions of state space. Less uniform energy distributions, which have relatively many macroscopic structures, correspond to smaller regions of state space. Systems featuring only Boltzmann Brains, undergoing experiences as of galaxies which are not actually there, have relatively few macroscopic structures – they only have the sorts of macroscopic structures which Boltzmann Brains have. Systems featuring ordinary observers, in galaxies which are actually there, have relatively many macroscopic structures – they have the sorts of macroscopic structures which ordinary observers, in ordinary galaxies, have. Therefore, regions of state space corresponding to Boltzmann Brain macrostates are far larger than regions of state space corresponding to ordinary observer macrostates. In addition to all this, the second law of thermodynamics implies that entropy either increases or stays the same with high probability, but not with probability one. So over long enough time scales, with probability approaching one, entropy eventually decreases. And that permits

---

[10]One important complication, for all these arguments, relates to the measure problem in cosmology. Very roughly, this is the problem of how to define a probability measure over events in the unboundedly large universes often discussed in Boltzmann Brain scenarios. For unbounded universes will contain infinitely many Boltzmann Brains and infinitely many ordinary observers – and different ways of partitioning those infinite collections up, using probability measures, generate different probabilities for the likelihood of any given individual being a Boltzmann Brain rather than an ordinary observer. For arguments in favor of using the stationary measure when assigning those probabilities, see (Linde et al., 2009). For empirical arguments against this choice of measure, see (Simone et al., 2010), who also argue that a different measure—the scale-factor cutoff measure—might not lead to a proliferation of Boltzmann Brains. For arguments claiming that most all proposed measures risk generating Boltzmann Brains, see (Linde & Noorbala, 2010).

a variety of different mathematical and physical arguments—based on Poincaré recurrence, ergodicity, typicality, and more—for the claim that the universe's microstate passes through state space regions corresponding to macrostates featuring only Boltzmann Brains far more often than it passes through state space regions corresponding to macrostates featuring only ordinary observers: since again, the former regions are so much bigger than the latter regions. Hence the existence of Boltzmann Brains.

2.2   A Precise Formulation of the Problem

With all that as background, here is the Boltzmann Brain problem in more detail. The following two postulates do a reasonably good job of codifying the many different formulations of the problem in the literature.[11]

Cosmological Inference
Given certain cosmological, thermodynamic, and statistical mechanical models, which are reasonably well-supported by the currently available scientific evidence, you should have extremely high credence that far more individuals undergoing experiences subjectively indistinguishable from yours are Boltzmann Brains rather than ordinary observers.

Generic Inference
If you should have extremely high credence that far more individuals undergoing experiences subjectively indistinguishable from yours are Boltzmann Brains rather than ordinary observers, then you should have extremely high credence in being a Boltzmann Brain.

Cosmological Inference connects scientific evidence to a normative claim about your credences: given the evidence, at least some reasonable models imply that you should have high credence in theories suggesting that among everyone having experiences subjectively indistinguishable from yours, far more are Boltzmann

---

[11]For similar formulations, see (Avni, 2023, p. 961; Dogramaci & Schoenfield, 2025, p. 3).

Brains than ordinary observers. Generic Inference connects one normative claim to another: if you should be confident that most instances of your subjective experiences, among all individuals, are had by a Boltzmann Brain rather than an ordinary observer, then you should be confident in being a Boltzmann Brain – since you are, presumably, generic among the individuals whose subjective experiences are indistinguishable from yours.

The problem posed by Boltzmann Brains is different from the problem which the simulation argument poses. For the theories which support the Boltzmann Brain problem enjoy more empirical support, have survived more cross-checks against competing hypotheses, and generate more novel predictions which additional experimentation continues to bear out, than the theories which support the simulation argument. In the Boltzmann Brain problem, the relevant probabilities are often taken to be derivable from classical and quantum versions of statistical mechanics, the Lambda-CDM model, general relativistic theories with de Sitter vacuum solutions, and more. These theories, and the parameters and laws and models which comprise them, have been stringently tested through impressively precise measurements: for example, of the Hubble constant, the mass density, the cosmological constant, and the deceleration parameter (Riess et al., 1998); and of the baryon density, the matter fluctuation amplitude, baryon acoustic oscillation, neutrino mass, and the universe's curvature (Aghanim et al., 2020). And these theories predict, imply, and are used to account for, a massive amount of empirical data which several different scientific fields have gathered. In the simulation argument, in contrast, the relevant probabilities are derived from plausibility arguments about what seems likely to us. For instance, the theories use counting, intuitive judgments, and Bayesian calculations based on these, to estimate that certain proportions have certain values (Bostrom, 2003, 2005; Chalmers, 2022). Perhaps those proportions do indeed represent credences which we are rationally required to have – that may well be right. Regardless, the principles used to estimate those proportions are not as empirically supported as the principles used to estimate the probabilities for Boltzmann Brains; nor have the former principles been used to account for large amounts of empirical data in this-or-that field of science. So the problem posed by Boltzmann Brains is, along this particular dimension at least, based on theories and calculations which are more secure than

the theories and calculations that the simulation argument features.

An important qualification: this is not to say that the theories used to argue for Boltzmann Brains have been definitively established, to the extent that other contemporary physical theories—like various aspects of the standard model of particle physics, for instance—have been. Some argue that Boltzmann Brains do not nucleate from the de Sitter vacuum in the way usually assumed (Boddy et al., 2017). Others argue that, given certain measure proposals for eternal inflation, small black hole nucleation can help avoid Boltzmann Brain domination (Olum et al., 2021). In addition, there are many mysteries in contemporary cosmology—concerning dark energy, say—which have not been addressed, and whose resolution might solve the Boltzmann Brain problem. Assuming that the cosmic horizon encloses an appropriately finite state space, perhaps because of certain features of dark energy, and so for instance the Lambda-CDM model really does provide an accurate description of the universe and continues to do so far into the future, the Boltzmann Brain problem arises. But these assumptions are quite strong, and reasonably contestable. In other words, the empirical data support a cosmological framework which is broader than the particular theories, with their certain specific assumptions, that generate the Boltzmann Brain problem. Specific assumptions about dark energy and the universe's Lambda-CDM structure, say—the assumptions needed to argue that Boltzmann Brains far outnumber ordinary observers—are not nearly as well-supported as the more general background cosmological framework is.[12]

There are roughly three kinds of response to the formulation of the Boltzmann Brains problem just given. First, provide justification for rejecting Cosmological Inference. Second, provide justification for rejecting Generic Inference. Third, accept the troubling conclusion of the problem: given the evidence, you should have high confidence in being a Boltzmann Brain. The rest of the paper reviews each of these.

---

[12]Thanks to David Wallace here.

3   Rejecting Cosmological Inference

One response to the Boltzmann Brain problem rejects Cosmological Inference: given the relevant models and evidence, you need not have extremely high credence that among individuals with experiences just like yours, most are Boltzmann Brains. In this section, I summarize two responses along these lines. To start, in Section 3.1, I explain how the Boltzmann Brain problem often gets described in two different ways, corresponding to two different views about what a cognitively unstable theory—or a self-undermining theory—is. In Section 3.2, I present a response to the problem which constrains the prior credences that agents should have. In Section 3.3, I present a response to the problem which invokes a distinction between two different kinds of evidence. Both responses appeal to some version of the claim that theories which proliferate Boltzmann Brains are cognitively unstable or self-undermining.

3.1   Preliminaries

In the literature on the Boltzmann Brain problem, the terms 'cognitively unstable' and 'self-undermining' are often used interchangeably. But another use of those terms, which does not treat them as synonymous, respects an important distinction in how different authors think about the Boltzmann Brain problem. That other use—and the corresponding distinction—is worth keeping in mind. Doing so helps clarify exactly how certain solutions to the Boltzmann Brain problem work, and exactly what kinds of issues those solutions address.

To illustrate the distinction, here is Elga's description of why theories which proliferate Boltzmann Brains might be cognitively unstable (2025, pp. 132-133). If you become highly confident in being a Boltzmann Brain, then you should think that your apparent evidence is non-veridical: you did not actually gather data about how gases behave, how material bodies interact with each other, and so on; the experience as of gathering that data resulted from a random fluctuation, rather than physical interactions with real gases and bodies. So your high confidence in being a Boltzmann Brain is unwarranted. And so you should have high confidence

in not being a Boltzmann Brain.[13] But if you become highly confident in not being a Boltzmann Brain, then you should think that your apparent evidence is veridical after all: as an ordinary observer, you really did gather data about how gases behave, how material bodies interact with each other, and so on. That evidence, however, supports theories which proliferate Boltzmann Brains. So your high confidence in not being a Boltzmann Brain is unwarranted. And so you should have high confidence in being a Boltzmann Brain; you are back to where you started.

So when it comes to cognitively unstable theories—in the present sense of 'cognitively unstable'—rationality seems to recommend incoherent credences. Given high credence in a cognitively unstable theory, it follows that in order to be rational, you should actually have low credence in that theory. And given low credence in a cognitively unstable theory, it follows that in order to be rational, you should actually have high credence in that theory. Put another way, if rationality allows for credence assignments to cognitively unstable theories at all, then the required assignments are incoherent: if the recommendation is high credence, then your credences should really be low; and if the recommendation is low credence; then your credences should really be high. But of course, in fact, rationality makes no such demands: no legitimate theory of rationality requires agents to be incoherent. So in cases of cognitive instability, the proper conclusion to draw is that rationality does not allow for assigning credences at all. You cannot have credences in cognitively unstable theories, on pain of incoherence.

In contrast, consider the following description—based on comments due to Albert (2000) and Wallace (2023)—of why theories which proliferate Boltzmann Brains are self-undermining. Once again, if you become highly confident in being a Boltzmann Brain, then you should think that your apparent evidence is non-veridical: again, you did not actually gather data on real gases and real material bodies. So again, your high confidence in being a Boltzmann Brain is unwarranted. But it need not follow, from this, that you should have high confidence in not being a Boltzmann Brain. And even if you do have high confidence in that, you need

---

[13]There are many reasons why at this juncture, your confidence in not being a Boltzmann Brain should be high rather than, say, middling. One reason: perhaps antecedently, you should have high confidence that you are not in a skeptical scenario.

not conclude that your apparent evidence is veridical. And most importantly, even if you did draw that conclusion, it does not follow that the purportedly veridical evidence supports theories which proliferate Boltzmann brains. That cannot follow: such theories undermine their own empirical base (Myrvold, 2016, p. 584); such theories undermine belief in themselves (Albert, 2000, p. 116). Or as Carroll puts it, we cannot rationally believe both that such theories are true and that we have good reason for that belief (2021, p. 7).

So when it comes to self-undermining theories—in the present sense of 'self-undermining'—rationality seems to recommend not believing in those theories. For in order to rationally believe that such theories are true, in order to have high rational credences in such theories, you must also have good reason for those beliefs. And you cannot have those good reasons while also holding onto your high rational credences. So rationality requires setting your credences in self-undermining theories to something quite low.

Put this way, the difference between cognitively unstable theories and self-undermining theories is clear. The rational response to cognitively unstable theories is to refrain from any credence assignment all. That is what rationality demands – that is the only way out of the incoherent circle sketched earlier. The rational response to self-undermining theories, in contrast, is to assign low credences. That is what rationality demands – that is the best response to the fact that you cannot rationally believe such theories and also have good reasons for that belief.

3.2   Constraining Rational Priors

Now let us consider responses to the Boltzmann Brain problem which reject Cosmological Inference; as will become clear, on my preferred reading, these responses engage most directly with the view that theories proliferating Boltzmann Brains are self-undermining in the sense just defined. In this section, let us consider Carroll's view that theories which proliferate Boltzmann Brains are self-undermining (2021, pp. 16-17). For if you are a Boltzmann Brain, then as just explained, none of the experiments which you take yourself to have performed—

and that provided the empirical basis for positing the theory which implies that Boltzmann Brains exist—ever actually occurred. Therefore, Cosmological Inference should be rejected. It is not the case that given the scientific models and evidence, you should be extremely confident that Boltzmann Brains far outnumber ordinary observers, and that many Boltzmann Brains' experiences are subjectively indistinguishable from yours.

To justify this rejection of Cosmological Inference, Carroll proposes a positive account of what our credences in self-undermining theories should be. Specifically, Carroll recommends setting our priors in self-undermining theories to astronomically small values (2021, p. 17): this ensures that given the scientific evidence, it is not the case that you should have extremely high credence in the proliferation of Boltzmann Brains. So self-undermining theories need not be ruled out by gathering evidence – for instance, by doing empirical experiments in an attempt to show that we are not Boltzmann Brains. Rather, self-undermining theories should be discarded from consideration before we begin gathering evidence at all, by adopting sufficiently low priors.

Against this, several argue that we should allow for having high credence in self-undermining theories, at least sometimes. Adapting a case due to Kotzen, suppose you have the experience as of ants arranging themselves to form the words "The Matrix has you" (2021, p. 26). Let the 'Matrix Hypothesis' be the theory that all your experiences are caused by electrical stimulations in The Matrix, and so your experiences are almost entirely non-veridical. Then the experience as of the ants is unreliable, since in fact the ants do not exist. So the Matrix Hypothesis is a self-undermining theory. But it would be unreasonable to dismiss the Matrix Hypothesis on that basis. It would be unreasonable to insist that your prior in the Matrix Hypothesis be so small as to prevent your credence, in the Matrix Hypothesis, from increasing significantly. If this is correct, then contrary to Carroll, prior credences in self-undermining theories need not be astronomically small: such theories need not be discarded from consideration before we begin gathering evidence.[14]

---

[14]For somewhat analogous criticisms of Carroll's rejection of Cosmological Inference, see (Avni, 2023, pp. 962-963). See also Dogramaci's approach to the Boltzmann Brain problem, which invokes the notion of cognitive stability, but by way of arguing against Generic Inference instead of Cosmological Inference (2020, pp. 3720-3721).

3.3   Two Kinds of Evidence

Wallace provides another response to the Boltzmann Brain problem which amounts to rejecting Cosmological Inference (2023). Unlike Carroll's response, however, Wallace's approach need not endorse the view that in order to be rational, agents must assign astronomically low prior credences to self-undermining theories. Instead, Wallace proposes a distinction between two different kinds of evidence, and on the basis of that distinction, argues that self-undermining theories should be assigned low conditional credence.

The distinction concerns what Wallace calls 'primary evidence' and 'proximal evidence' (2023, p. 296). The primary evidence for a scientific claim is a conjunction of many statements like the following: the second law of thermodynamics passed thus-and-so test from the mid-1800s; the second law of thermodynamics passed a more recent test; calculations based on the Boltzmann equation for gas diffusion correctly predict such-and-such experimental result; this ice cube melted over the course of ten minutes; and so on. So the primary evidence for a claim consists of various experiments and natural phenomena, from many different periods of history, known to most of us only through testimony. The proximal evidence for a scientific claim, in contrast, is a conjunction of statements about much more direct and immediate evidential experiences, like the following: I remember reading that the second law of thermodynamics passed thus-and-so test from the mid-1800s; I remember David Wallace telling me that the second law of thermodynamics passed a more recent test; I remember watching a video about correct predictions generated by calculations based on the Boltzmann equation for gas diffusion; I see this ice cube melting; and so on. So the proximal evidence for a claim consists of records, testimony, and memories that—in normal circumstances—provide good reason for believing the primary evidence.

The distinction between primary evidence and proximal evidence is attractive for several reasons. Perhaps most importantly for present purposes: the distinction supports an elegant account of what it is for a theory to be self-undermining. According to the account, theory T is self-undermining, relative to proximal evidence E and primary evidence H—and relative to a given agent's probabilistic credence function Pr—if and only if $Pr(H \mid T\&E)$ is extremely

low (Wallace, 2023, p. 297). In other words, T is self-undermining—relative to your proximal evidence, primary evidence, and credences—just in case you are highly confident that the primary evidence, conditional on both the theory being true and the proximal evidence holding, never actually happened. It follows that theories which proliferate Boltzmann Brains are self-undermining relative to the relevant proximal evidence, primary evidence, and credences.

As Wallace shows, given some assumptions, this implies that you have no good reason to assign high credence to a self-undermining theory T (2023, pp. 297-298). Slightly more precisely, suppose that T is self-undermining relative to proximal evidence E, primary evidence H, and probabilistic credence function Pr. So $Pr(H \mid T\&E) \ll 1$. In addition, suppose that Pr assigns low credence to skeptical scenarios in which the proximal evidence holds but the primary evidence does not – that is, to situations in which your memories of reading about tests of the second law, your memories of hearing about such tests, your visual perception as of a melting ice cube, and so on, are non-veridical. So $Pr(\neg H \mid E) \ll 1$. Then as a straightforward calculation shows, the prior credence in T, conditional on E, is extremely small too. That is, $Pr(T \mid E) \ll 1$.

This supports the rejection of Cosmological Inference. By Wallace's account of what self-undermining is, theories which proliferate Boltzmann Brains are self-undermining relative to the relevant proximal evidence, primary evidence, and credences. Therefore, the rational credence to have in such theories, conditional on the relevant proximal evidence, is extremely low. So it is not the case that given the cosmological, thermodynamic, and statistical mechanical models and evidence, you should assign high credence to far more individuals having experiences subjectively indistinguishable from yours being Boltzmann Brains rather than ordinary observers.

In response to Wallace's argument, one might deny (i) the claim that theories which proliferate Boltzmann Brains are self-undermining, or (ii) the assumption that you should assign low credence to skeptical scenarios where the proximal evidence holds but the primary evidence does not. In support of denying (i), note that this claim relies on an account of self-undermining theories which itself presupposes a distinction between primary evidence and proximal evidence; and it might be difficult to draw that distinction cleanly. For that distinction

is somewhat reminiscent of the distinction between observation sentences and theoretical sentences. And just as one might think that observation sentences are theory-laden, one might think that proximal evidence is 'primary evidence'-laden, in the sense that what counts as proximal evidence rather than primary evidence can vary across different theoretical contexts. In support of denying (ii), one might claim that (ii) is, though perhaps plausible in certain dialectical contexts, question-begging in the dialectical context at issue here. For during debates over whether to accept the troubling conclusion of the Boltzmann Brain problem, claims about the relationship between proximal evidence and primary evidence are precisely the claims under dispute. To assume that the proximal evidence is not misleading as to the primary evidence—that is, to endorse (ii)—is to take for granted that we are not in the sort of skeptical scenario which the possibility of Boltzmann Brains raises (Dogramaci & Schoenfield, 2025, p. 21).

4   Rejecting Generic Inference

Another response to the Boltzmann Brain problem—which tends to be the most popular response in the literature—rejects Generic Inference: even if you should be confident that more individuals having experiences like yours are Boltzmann Brains than ordinary observers, you need not be confident that you are a Boltzmann Brain. In this section, I summarize two responses along these lines. The first, discussed in Section 4.1 invokes the Past Hypothesis, which makes a claim about what the universe's initial macrostate was like. The second, discussed in Section 4.2 concerns the sorts of experiences that Boltzmann Brains are likely to have.

4.1   The Past Hypothesis

The Past Hypothesis is a law which states that initially, the universe had extremely low entropy (Albert, 2000, pp. 95-96). As a matter of nomological necessity, the initial microstate of the universe belonged to a state space region

of extremely small size. Standard cosmological experiments and inferences tell us what this region's low-entropy, highly condensed, big-bang sort of macrostate is like (Albert, 2000, p. 96; Feynman, 1965, p. 116). The Past Hypothesis is simply a law which says that the universe's initial microstate belonged to the region corresponding to that macrostate which cosmology describes.

Before discussing the plausibility of the Past Hypothesis, let us see how it can be used to justify rejecting Generic Inference. Suppose that as the Past Hypothesis implies, about 13.8 billion years ago, the universe's entropy was miniscule. Then with extremely high probability, there are no Boltzmann Brains at all. There has not been enough time for the probability of the relevant fluctuations, producing Boltzmann Brains, to have occurred. With probability more-or-less equal to one, the universe has undergone standard thermodynamic evolution throughout the last 13.8 billion years, its entropy increasing as it passes through macrostates bringing it closer and closer to equilibrium. Now, it is still true that throughout the entire history of the universe—including the universe's entire future—the vast majority of individuals having experiences like yours are Boltzmann Brains rather than ordinary observers. You should still be confident in that, for all the reasons discussed in Section 2. But since you know the Past Hypothesis, you should also be confident that all those Boltzmann Brains, with experiences subjectively indistinguishable from yours, exist in the far future. So you should be confident in being an ordinary observer.

There is a reasonable concern about all this: positing the Past Hypothesis merely to avoid the Boltzmann Brain problem might seem cheap and unmotivated. The Past Hypothesis does indeed imply that with high probability, we are all ordinary observers. But are there any other reasons, independent of the solution to the Boltzmann Brain problem which it supports, to posit the Past Hypothesis? If not, then solving the Boltzmann Brain problem by positing the Past Hypothesis seems uncompelling.
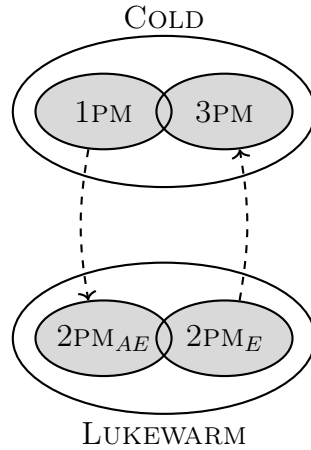
As it turns out, there are plenty of other reasons to posit the Past Hypothesis. Perhaps most importantly, the Past Hypothesis provides a defense against a well-known reversibility objection to the interpretation of the second law given earlier.[15]

---

[15]Other important reasons concern accounting for the arrow of time (Chen, 2023; Loewer, 2024; North, 2011).

To explain how the reversibility objection works, consider a system consisting of a cup of coffee together with its immediate environment. Take two macrostates of that system: the macrostate corresponding to the coffee being lukewarm, and the macrostate corresponding to the coffee being cold. Let LUKEWARM, and COLD be the regions of the system's state space corresponding to these two macrostates; note that in what follows, for convenience, I also use 'LUKEWARM' and 'COLD' to reference the corresponding macrostates themselves. Suppose you come across a lukewarm cup of coffee: so its microstate is in LUKEWARM. How confident should you be that an hour before, the system's microstate was in the region COLD – in other words, that over the course of the past hour, the coffee has gotten warmer rather than colder?

Without a posit analogous to the Past Hypothesis, an argument shows that you should be exactly as confident that the system's microstate was in COLD, an hour ago, as that the system's microstate will evolve to COLD in an hour.[16] You should be as confident that the coffee has gotten warmer over the course of the past hour, in other words, as you are that the coffee will get colder over the course of the hour to come. To see why, suppose we observe the coffee at 2pm, and we find that the coffee is lukewarm: so our total evidence is that (i) the system's microstate is in the LUKEWARM macrostate at 2pm, and of course (ii) various dynamical laws hold. Let 2PM_E be the subregion of LUKEWARM consisting of microstates lying on entropic trajectories that pass through LUKEWARM at 2pm and pass through COLD at 3pm. Let 2PM_{AE} be the subregion of LUKEWARM consisting of microstates lying on anti-entropic trajectories that pass through COLD at 1pm and pass through LUKEWARM at 2pm. Then the size of the region 2PM_E, as quantified using the standard measure for systems like these, is exactly the same as the size of the region 2PM_{AE}, as quantified using that measure. The picture below can be used to explain why.

---

[16]For versions of this argument, see (Loschmidt, 1876). For somewhat similar objections to the Past Hypothesis based on recurrence, see (Zermelo, 1896).

In this picture, the large unshaded ovals represent the macrostates LUKEWARM and COLD. The lower two shaded ovals represent the subregions of LUKEWARM corresponding to 2PM_{AE} and 2PM_E. The upper shaded oval 3PM represents the microstates in COLD to which the microstates in 2PM_E evolve from 2pm to 3pm, and the corresponding dashed arrow represents that evolution. The upper shaded oval 1PM represents the microstates in COLD which evolve to microstates in 2PM_{AE} from 1pm to 2pm, and the corresponding dashed arrow represents that evolution. The size of 3PM equals the size of 2PM_E, and the size of 1PM equals the size of 2PM_{AE}: this follows from the fact that the dynamical laws preserve the sizes of sets when size is quantified using the standard measure mentioned earlier. In addition, the size of 3PM equals the size of 1PM: this follows from, among other things,[17] (i) 1PM can be obtained by reversing the velocities of the microstates in 3PM,[18] (ii) the sizes of regions are invariant under this velocity-reversing transformation,[19] and (iii) the dynamical laws are time-reversal invariant.[20] Therefore, 2PM_{AE} and 2PM_E have the same size. Since by standard typicality postulates,

---

[17]For additional discussion of this argument, which explores these facts and other important assumptions in more detail than I can here, see (Albert, 2000, pp. 71-76).

[18]This velocity-reversing transformation, note, preserves macrostates: that is, if a microstate belongs to a region corresponding to a given macrostate, then reversing the velocities of the particles in that microstate generates another microstate belonging to that same region.

[19]That is, each region X has the same size as any region obtained from X by reversing the velocities of X's microstates.

[20]The time-reversal invariance of the dynamical laws ensures that since each microstate in 3PM was in 2PM_E an hour ago, the microstates in 1PM—obtained by the velocity-reversing transformation—evolve to LUKEWARM in an hour.

your degree of confidence that the system's microstate belongs to a subregion of LUKEWARM should equal the size of that subregion, it follows that your confidence in the coffee having been colder, an hour ago, should equal your confidence that the coffee will become colder in the next hour.

That is bizarre. It contradicts the simple fact that whenever we encounter lukewarm cups of coffee in the world, they basically always were warmer in the past. This is one of the reversibility objections to the statistical mechanical approach to entropy and the second law from earlier: that interpretation, based on measures which assign sizes to regions of state space, generates predictions for how coffee cups behave—and specifically, retrodictions[21] for how coffee cups have behaved—which our experiences contradict.

The Past Hypothesis can be used to defend the statistical mechanical approach against all this. For the Past Hypothesis rules out many of the entropy-decreasing trajectories that pass through 1PM: that is, the trajectories which, while currently in LUKEWARM, passed through COLD an hour ago. For the Past Hypothesis says that initially, the universe had extremely low entropy. And as it turns out, the vast majority of trajectories which both have an initial microstate in the region corresponding to that initial low entropy macrostate, and also pass through LUKEWARM at 2pm, were not in COLD an hour ago. Slightly more technically: even though 2PM_{AE} and 2PM_E have the same size, $2\text{PM}_{AE} \cap$ PH LAW – the subregion of 2PM_{AE} consisting of microstates lying on trajectories which originated in the universe's low-entropy initial macrostate – is far, far smaller than $2\text{PM}_E \cap$ PH LAW – the subregion of 2PM_E consisting of microstates lying on trajectories which originated in the universe's low-entropy initial macrostate. For far fewer microstates evolved from the universe's initial macrostate to COLD and then to LUKEWARM, than evolved from the universe's initial macrostate to LUKEWARM without passing through COLD first.[22]

---

[21]A retrodiction is like a prediction except that while predictions involve inferences about what the future will be like, retrodictions involve inferences about what the past was like.

[22]Put in terms of conditionalization and total evidence, the point is this. Let μ be the measure, mentioned earlier, which assigns sizes to subsets of state space. Let DYNAMICAL LAWS be the set of all microstates at which the dynamical laws hold: so trivially, DYNAMICAL LAWS is the set of all microstates whatsoever; I mention this set merely to make clear that these laws are part of our total evidence. Then $\mu(2\text{PM}_{AE}|\text{LUKEWARM} \cap \text{DYNAMICAL LAWS}) = \mu(2\text{PM}_E|\text{LUKEWARM} \cap \text{DYNAMICAL LAWS})$ but $\mu(2\text{PM}_{AE}|\text{LUKEWARM} \cap \text{DYNAMICAL LAWS} \cap$

In short, by positing the Past Hypothesis along with the dynamical laws—and other claims about measures quantifying the sizes of sets—we get a package of principles which yields the correct results about how coffee cools rather than warms, how gasses expand rather than contract, how ice cubes melt rather than get colder, and so on. Without the Past Hypothesis, the package of principles which we use to calculate phenomena like that—consisting of the dynamical laws, and claims about probability measures—gives obviously wrong results. But that is to be expected: it makes sense that if you neglect to take account of all the physical laws, your predictions and retrodictions will go awry. So we should indeed adopt the Past Hypothesis as part of our total nomological package.[23]

So there is nothing cheap and unmotivated about positing the Past Hypothesis to reject Generic Inference, and in so doing, solve the Boltzmann Brain problem. In order to get even the simplest predictions and retrodictions about thermodynamics right, we need the Past Hypothesis. Without it, our physical theory does not even correctly imply that with high probability, coffee cools rather than warms. With it, our physical theory does indeed have that implication.

In fact, once we adopt the Past Hypothesis, Generic Inference is the posit which looks problematically cheap and unmotivated. The original arguments for Boltzmann Brains with experiences like yours being far, far more common than ordinary observers with experiences like yours, invoked the dynamical laws, an account of entropy, and a certain interpretation of the second law of thermodynamics. But if the Past Hypothesis is a law, then those arguments were based on a proper subportion of our total physical theory of the world. Those arguments left part of the total theory out. And it is unsurprising that arguments based on an incomplete portion of the total physical theory would draw strange conclusions. Getting worried about Boltzmann Brains, because of arguments which neglect

PH LAW) $\ll \mu(2\text{PM}_E|\text{LUKEWARM} \cap \text{DYNAMICAL LAWS} \cap \text{PH LAW})$. So if the Past Hypothesis is in your total evidence, then the problem is avoided: assuming that you credences should match the typicality measure μ, it follows that your credence in 2PM_{AE}, conditional on your total evidence, is far lower than your credence in 2PM_E conditional on that evidence. Therefore, you should be much more confident that the coffee will be colder in an hour than that the coffee was colder an hour ago.

[23]As Albert puts it, our grounds for believing the Past Hypothesis "have to do with the fact that the proposition that the universe came into being in an enormously low-entropy macrocondition turns out to be enormously helpful in making an enormous variety of particular empirical predictions" (2000, pp. 93-94).

the Past Hypothesis, is like getting worried that nothing ever moves, because of arguments which neglect the dynamical laws. Perhaps the worrisome conclusions do indeed follow, but so what? The worrisome conclusions do not follow when all the laws are taken into account.

An aside: in my view, this is one place where it may well matter that the Past Hypothesis—when wielded as a response to the Boltzmann Brian problem—be treated as a law. If the Past Hypothesis were a nomologically contingent fact, then it would not be the case that the original arguments for Boltzmann Brains left part of the total physical theory of the world out. And so those arguments would be more compelling: the worrisome conclusions would indeed follow when all the laws are taken into account.[24]

There are other potential problems with the Past Hypothesis, however. Some argue that the Past Hypothesis is not even well-formulated in general relativistic spacetime (Earman, 2006). Others argue that the Past Hypothesis cannot be used to explain the entropic evolution of local isolated systems, much smaller than the size of the universe – including, perhaps, the coffee mentioned earlier (Winsberg, 2004). These arguments, if successful, would undermine any response to the Boltzmann Brain problem which presupposes the Past Hypothesis.

4.2   Ordered and Disordered Experiences

Another justification for rejecting Generic Inference appeals to the likelihoods of having certain sorts of experiences, or being in certain sorts of circumstances. Put intuitively, the idea is this. Consider the following question: does your total evidence raise the probability that you are a Boltzmann Brain, or the probability that you are an ordinary observer, by more? As some calculations based on Bayes' theorem show, given a reasonable anti-skeptical assumption, the following holds: (i) your total evidence raises the probability of being a Boltzmann Brain more than it raises the probability of being an ordinary observer, if and only if (ii) the probability of that total evidence, conditional on being a Boltz-

---

[24]For arguments against the claim that the Past Hypothesis needs to be treated as a law, in order to reap various benefits, see (Lazarovici & Reichert, 2020).

mann Brain, is greater than the probability of that total evidence conditional on being an ordinary observer. Now, pre-theoretically, it seems that (ii) is not true. For Boltzmann Brains have a far wider range of experiences than ordinary observers have – the vast majority of those experiences are disordered, incoherent, non-continuous and generally unlike our own (Dogramaci & Schoenfield, 2025; Kotzen, 2021; Page, 2024a). So the probability of having total evidence generated by the kind of ordered, coherent, continuous experiences that you in fact have, conditional on your being a Boltzmann Brain, is far less than the probability of having that total evidence conditional on your being an ordinary observer. Therefore, (i) fails: your total evidence raises the probability of being an ordinary observer more than it raises the probability of being a Boltzmann Brain. And so Generic Inference is false.

Here is one formalization of this argument, which follows the particularly clear presentation in (Dogramaci & Schoenfield, 2025, pp. 4-9). Let Pr be your credence function, let Y_b be the proposition that you are a Boltzmann Brain, let Y_o be the proposition that you are an ordinary observer, and let E be your total evidence. So $Pr(Y_b \mid E)$ and $Pr(Y_o \mid E)$ are your posterior credences in being a Boltzmann Brain and in being an ordinary observer, respectively, conditional on the evidence which you have. In order for your total evidence to raise the probability of being a Boltzmann Brain more than it raises the probability of being an ordinary observer, $Pr(Y_b \mid E)$ must be greater than $Pr(Y_o \mid E)$. In other words,

$$\frac{Pr(Y_b \mid E)}{Pr(Y_o \mid E)} > 1 \tag{1}$$

Bayes' theorem implies that

$$\frac{Pr(Y_b \mid E)}{Pr(Y_o \mid E)} = \frac{Pr(Y_b)}{Pr(Y_o)} \frac{Pr(E \mid Y_b)}{Pr(E \mid Y_o)} \tag{2}$$

Adopt the anti-skeptical assumption that your prior credence in being a Boltzmann Brain is not greater than your prior credence in being an ordinary observer. In other words,

$$Pr(Y_b) \leqslant Pr(Y_o) \tag{3}$$

Then (3) and (2) imply that (1) holds if and only if

$$Pr(E \mid Y_b) > Pr(E \mid Y_o) \tag{4}$$

holds; that is, if and only if the likelihood of your evidence, conditional on your being a Boltzmann Brain, is greater than the likelihood of your evidence conditional on your being an ordinary observer.

But (4) is implausible. For in English, it says that the probability of having the total evidence which you in fact have, conditional on being a Boltzmann Brain, is greater than the probability of having that total evidence conditional on being an ordinary observer. And that seems wrong, because Boltzmann Brains are capable of having a far wider and more disunified range of experiences than ordinary observers are capable of having. As Kotzen puts it, even if you did fluctuate into existence, it is overwhelmingly improbable that you would have the sorts of ordered, coherent experiences which you in fact have – since the vast majority of conscious Boltzmann Brains have wildly disordered and incoherent experiences (2021, pp. 29-30). As Page puts it, the fraction of Boltzmann Brain observations that are ordered is much less than the fraction of ordinary observer observations that are ordered, since so many Boltzmann Brains have disordered observations (2024b, p. 1; 2024a, pp. 61-62). As Dogramaci and Schoenfield put it, since Boltzmann Brains are the product of random processes, the range of possible evidential experiences which a Boltzmann Brain could have is far, far greater than the range of possible evidential experiences which an ordinary observer could have – while ordinary observers' experiences must be reasonably coherent, Boltzmann Brains' experiences are totally unconstrained (2025, p. 8). So (4) fails. Therefore, (1) fails as well. And so your total evidence does not raise the probability of being a Boltzmann Brain more than it raises the probability of being an ordinary observer.

As Saad argues, this rejection of (4) faces problems (Saad, 2024). The notion of experiential disorder, and the connection between that notion and the notion of total evidence, is somewhat intuitive but fairly imprecise (Saad, 2024, p. 5). It is unclear what counts as a disordered, incoherent, non-continuous experience, and what counts as an ordered, coherent, continuous experience. It is also unclear how

these sorts of experiences bear on the likelihoods of having certain sorts of total evidence. And absent accounts of all this, the proposed rejection of (4) might seem unconvincing.[25]

Another potential problem with rejecting (4) relates to Boltzmann Brains with ordered, coherent, continuous experiences. As discussed earlier, there are far more Boltzmann Brains with experiences subjectively indistinguishable from yours than there are ordinary observers with experiences subjectively indistinguishable from yours. All of these beings have ordered, coherent, and continuous experiences – Boltzmann Brains and ordinary observers alike. And this fact can be used to formulate another version of the Boltzmann Brains problem.

To see how, let Y_b' be the proposition that you are one of the Boltzmann Brains with ordered, coherent, and continuous experiences which are subjectively indistinguishable from yours. Then plausibly, $Pr(E \mid Y_b')$ is fairly high. The probability of your total evidence E obtaining, given that you are a Boltzmann Brain having whatever ordered, coherent, and continuous experiences you in fact have, is large: for on several reasonable conceptions of how your experiences and your total evidence relate, being an observer—either ordinary or Boltzmann Brain—with whatever experiences you have, makes whatever total evidence you have quite likely. In addition, plausibly, $Pr(E \mid Y_o)$ is fairly low. The probability of your total evidence E obtaining, conditional on your being some ordinary observer or other, is small: for many ordinary observers have experiences quite different from yours, and so on several reasonable conceptions of how your experiences and your total evidence relate, it follows that many observers have total evidence E' distinct from—in many cases, even mutually exclusive with—E. Therefore, $Pr(E \mid Y_b') \gg Pr(E \mid Y_o)$. And therefore, so long as

$$\frac{Pr(E \mid Y_b')}{Pr(E \mid Y_o)} > \frac{Pr(Y_o)}{Pr(Y_b')}$$

which seems plausible, it follows that

$$\frac{Pr(Y_b' \mid E)}{Pr(Y_o \mid E)} > 1 \qquad (5)$$

In other words, your total evidence raises the probability of being a Boltzmann Brain—specifically, one of the Boltzmann Brains with ordered, coherent, and continuous experiences which are subjectively indistinguishable from yours—more than it raises the probability of being an ordinary observer.

5 Accepting the Conclusion

One more response to the Boltzmann Brain problem is worth considering: simply accept both Cosmological Inference and Generic Inference. It follows that given the relevant scientific models and evidence, you should have extremely high credence in being a Boltzmann Brain. So the Boltzmann Brain problem is, in fact, no problem at all. Assuming that the relevant models do indeed describe our universe, the Boltzmann Brain problem is simply a striking fact with which we must reconcile ourselves.

Avni (2023) develops a version of this response. By way of preparation, make the following simplifying assumptions (Avni, 2023, pp. 960-961).

(1) Either Boltzmann Brains vastly outnumber ordinary observers, or ordinary observers vastly outnumber Boltzmann Brains; and you know this.

(2) Every individual is either a Boltzmann Brain or an ordinary observer; and you know this.

In addition, Avni defends the following principles (2023, pp. 962-969).

(3) Your credence in Boltzmann Brains vastly outnumbering ordinary observers, conditional on your being an ordinary observer, should be approximately one.

(4) Your credence in being a Boltzmann Brain, conditional on Boltzmann Brains vastly outnumbering ordinary observers, should be approximately one; as should your credence in being an ordinary observer, conditional on ordinary observers vastly outnumbering Boltzmann Brains.

Suitably formalized, principles (1)–(4) imply that your credence in being a Boltzmann Brain should be approximately one (Avni, 2023).

Avni provides several lines of support for (3) and (4). For instance, Avni argues that rejecting (3) – which is analogous to rejecting Cosmological Inference – is far too revisionary of standard scientific methodology (2023, p. 963). And Avni argues that rejecting (4) – which is analogous to rejecting Generic Inference – for instance by endorsing the sorts of views developed in Section 4, is either irrelevant or unconvincing for the issue at hand (2023, pp. 965-968). So ultimately, Avni concludes that if you have apparent memories as of receiving strong scientific evidence that Boltzmann Brains vastly outnumber ordinary observers, then you should indeed be confident in being a Boltzmann Brain (2023, p. 971).

There are ways to resist Avni's arguments. For instance, take Avni's criticisms of the rejection of the principle (4). Avni calls this principle 'Typicality', presumably because in Avni's estimation, it captures the following idea: in general, you should think that you are typical. So if most – that is, typical – agents are Boltzmann Brains, then you should think that you are a Boltzmann Brain. And if most – that is, typical – agents are ordinary observers, then you should think that you are an ordinary observer. In fact, Avni uses this kind of typicality reasoning to argue against responses to the Boltzmann Brain problem which attempt to reject (4) by invoking the Past Hypothesis. If Boltzmann Brains vastly outnumber ordinary observers, Avni claims, then it is still overwhelmingly more likely that any given brain—yours included—is a Boltzmann Brain; so the Past Hypothesis does not provide a reason to reject (4), or Generic Inference for that matter.

According to certain theories of typicality, however, the Past Hypothesis does indeed provide such a reason. My preferred theory of typicality, for instance, endorses what I call the 'Typical Principle': roughly, if a proposition typically obtains at the set of all physically possible worlds compatible with your evidence,

then you should believe that proposition (Wilhelm, in press-b, p. 5).[26] Now suppose that, as argued earlier, the Past Hypothesis is a law. So at all physically possible worlds compatible with your evidence, the Past Hypothesis holds. Then as argued in Section 4, at all physically possible worlds, the following proposition obtains: with extremely high probability, there are no Boltzmann Brains at all. So you should believe that with extremely high probability, there are no Boltzmann Brains. And so plausibly, your credence in anyone being a Boltzmann Brain at present, yourself included, should be extremely low. It follows that (4) is false.

Another interesting response to the Boltzmann Brain problem, somewhat similar in spirit to Avni's response, can be extracted from a view developed by Loew (2017). As Loew points out, in addition to supporting Boltzmann Brains, some cosmological and statistical mechanical models also support the view that each of us has many 'Boltzmann duplicates', where a Boltzmann duplicate of a person is an exact duplicate of that person as they were shortly before their death (2017, p. 763). Loew argues that on standard accounts of personal identity over time, your Boltzmann duplicates are you: more specifically, assuming that survival of death is possible at all, you will almost certainly survive your own death, because you are related to your Boltzmann duplicates in a way which ensures your survival (2017, p. 775). This is not to claim that you should be quite confident in being, right now, a Boltzmann Brain. Rather, this is just to claim that your connection to individuals very much like Boltzmann Brains is akin to your relationship to your past self. Those Boltzmann-Brain-like beings are, in that sense, parts of you.

If this is right, then perhaps the conclusion of the Boltzmann Brain problem—that you are, right now, a Boltzmann Brain—is not so troubling after all. Just as you stand in psychological continuity relations to your past self, you stand in psychological continuity relations to individuals which fluctuate into existence in the way that Boltzmann Brains do. And that might make it more palatable to think that you are probably, at present, a Boltzmann Brain.

---

[26]For more details of this theory, see (Wilhelm, 2022, 2025, in press-a). For alternative views of how typicality relates to the Boltzmann Brain problem, see (Hartle & Hertog, 2017; Hartle & Srednicki, 2007).

6 Conclusion

Boltzmann Brains raise a fascinating cluster of philosophical issues. Unlike other skeptical scenarios in the literature, the Boltzmann Brain problem is based on physical theories and calculations which enjoy a decent degree of empirical, scientific support. The problem reveals subtle connections which sometimes obtain between evidence and theory: for instance, how a theory can undermine the very evidence used to support it.

Responses to the Boltzmann Brain problem vary widely. Some respond by adopting constraints on rational priors. Others distinguish different kinds of evidence. Several propose new physical laws. Many claim that our experiences are orderly in a way that Boltzmann Brains' experiences are not. And some suggest accepting the problem's implication that, given the evidence, we are probably Boltzmann Brains.[27]

In my view, one of the most interesting features of the problem is the interaction it reveals between science and philosophy. Solving the problem requires engaging with both how confirmation works, in actual scientific practice, and also with theories of rational credence developed in philosophy. Is the Boltzmann Brain problem another place where ideas from physics, when developed with philosophical care, can solve puzzles in the literature on rationality – as solutions based on the Past Hypothesis suggest? Or is the Boltzmann Brain problem best addressed by appealing to claims about the requirements of rationality only – as appeals motivated by Bayesianism, based on likelihoods, suggest? And how does all this bear on the larger paradigm of using Bayesian confirmation theory to elucidate the relationship between theory and evidence in science? These are the sorts of questions which Boltzmann Brains raise and motivate.

---

[27]There are many other responses too, far more than I can summarize here. For responses based on Bohmian mechanics, see (Goldstein et al., 2025; Tumulka, 2022).

Acknowledgements

Thanks to David Albert, Shelly Goldstein, Dustin Lazarovici, Christian Löw, Rodi Tumulka, and David Wallace, for much helpful feedback and discussion.

References

Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., . . . Zonca, A. (2020). Planck 2018 results. *Astronomy and Astrophysics*, *641*, 1–67.

Albert, D. (2000). *Time and chance.* Cambridge, MA: Harvard University Press.

Albrecht, A. (2004). Cosmic inflation and the arrow of time. In J. D. Barrow, P. C. W. Davies, & C. L. Harper (Eds.), *Science and ultimate reality* (pp. 363–401). Cambridge: Cambridge University Press.

Albrecht, A., & Sorbo, L. (2004). Can the universe afford inflation? *Physical Review D*, *40*(063528), 1–10.

Avni, R. (2023). The boltzmann brains puzzle. *Noûs*, *57*, 958–972.

Boddy, K. K., Carroll, S. M., & Pollack, J. (2017). Why boltzmann brains do not fluctuate into existence from the de Sitter vacuum. In K. Chamcham, J. Silk, J. D. Barrow, & S. Saunders (Eds.), *The philosophy of cosmology* (pp. 228–240). New York, NY: Cambridge University Press.

Boltzmann, L. (2015). On the relationship between the second fundamental theorem of the mechanical theory of heat and probability calculations regarding the conditions for thermal equilibrium (K. Sharp & F. Matschinsky, Trans.). *Entropy*, *17*, 1971–2009. (Original work published 1877)

Bostrom, N. (2003). Are we living in a computer simulation? *The Philosophical Quarterly*, *53*(211), 243–255.

Bostrom, N. (2005). The simulation argument. *The Philosophical Quarterly*, *55*(218), 90–97.

Bousso, R., & Freivogel, B. (2007). A paradox in the global description of the multiverse. *Journal of High Energy Physics*, *06*, 1–7.

Carroll, S. M. (2010). *From eternity to here.* Oxford: Oneworld.

Carroll, S. M. (2021). Why boltzmann brains are bad. In S. Dasgupta, R. Dotan, & B. Weslake (Eds.), *Current controversies in philosophy of science* (pp. 7–20). New York, NY: Routledge.

Chalmers, D. J. (2022). *Reality+.* New York, NY: W W Norton.

Chen, E. K. (2023). The past hypothesis and the nature of physical laws. In B. Loewer, B. Weslake, & E. Winsberg (Eds.), *The probability map of the*

universe (pp. 204–248). Cambridge, MA: Harvard University Press.

Cotler, J., Hunter-Jones, N., & Ranard, D. (2022). Fluctuations of subsystem entropies at late times. *Physical Review A*, *105*(022416), 1–20.

Dogramaci, S. (2020). Does my total evidence support that i'm a boltzmann brain? *Philosophical Studies*, *177*, 3717–3723.

Dogramaci, S., & Schoenfield, M. (2025). Why i am not a boltzmann brain. *The Philosophical Review*, *134*(1), 1–33.

Dyson, L., Kleban, M., & Susskind, L. (2002). Disturbing implications of a cosmological constant. *Journal of High Energy Physics*, *10*, 1–22.

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

Eddington, A. S. (1931). The end of the world. *Nature (Supplement)*, *127*(3203), 447–453.

Elga, A. (2025). Boltzmann brains and cognitive instability. *Philosophy and Phenomenological Research*, *111*, 127–136.

Feynman, R. (1965). *The character of physical law.* Cambridge, MA: MIT Press.

Goldstein, S. (2001). Boltzmann's approach to statistical mechanics. In J. Bricmont, D. D. M. C. Galavotti, G. Ghirardi, F. Petruccione, & N. Zanghì (Eds.), *Chance in physics* (pp. 39–54). Heidelberg: Springer.

Goldstein, S., Struyve, W., & Tumulka, R. (2025). The bohmian approach to the problems of cosmological quantum fluctuations. *The British Journal for the Philosophy of Science*, *76*(4), 869–894.

Hartle, J., & Hertog, T. (2017). One bubble to rule them all. *Physical Review D*, *95*(123502), 1–15.

Hartle, J., & Srednicki, M. (2007). Are we typical? *Physical Review D*, *75*(123523), 1–6.

Kotzen, M. (2021). What follows from the possibility of boltzmann brains? In S. Dasgupta, R. Dotan, & B. Weslake (Eds.), *Current controversies in philosophy of science* (pp. 7–20). New York, NY: Routledge.

Lazarovici, D., & Reichert, P. (2015). Typicality, irreversibility and the status of macroscopic laws. *Erkenntnis*, *80*, 689–716.

Lazarovici, D., & Reichert, P. (2020). Arrow(s) of time without a past hypothesis. In V. Allori (Ed.), *Statistical mechanics and scientific explanation* (pp. 343–

386). Singapore: World Scientific.

Lebowitz, J. L. (1993a). Boltzmann's entropy and time's arrow. *Physics Today*, 32–38.

Lebowitz, J. L. (1993b). Macroscopic laws, microscopic dynamics, time's arrow and boltzmann's entropy. *Physical Review A*, *194*, 1–27.

Linde, A., & Noorbala, M. (2010). Measure problem for eternal and non-eternal inflation. *Journal of Cosmology and Astroparticle Physics*, *9*, 1–15.

Linde, A., Vanchurinc, V., & Winitzkic, S. (2009). Stationary measure in the multiverse. *Journal of Cosmology and Astroparticle Physics*, *1*, 1–22.

Loew, C. (2017). Boltzmannian immortality. *Erkenntnis*, *82*, 776–761.

Loewer, B. (2024). *Laws of nature and chances*. New York, NY: Oxford University Press.

Loschmidt, J. (1876). Über den zustand des wärmegleichgewichtes eines systems von körpern mit rücksicht auf die schwerkraft. i. *Sitzungsberichte der mathematisch-naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften*, *73*(2), 128–142.

Myrvold, W. (2016). Probabilities in statistical mechanics. In A. Hájek & C. Hitchcock (Eds.), *The oxford handbook of probability and philosophy* (pp. 573–600). Oxford: Oxford University Press.

North, J. (2011). Time in thermodynamics. In C. Callender (Ed.), *The oxford handbook of philosophy of time* (pp. 312–350). Oxford: Oxford University Press.

Norton, J. D. (2022). You are not a boltzmann brain. *International Journal of Quantum Foundations*, *8*, 190–194.

Olum, K. D., Upadhyay, P., & Vilenkin, A. (2021). Black holes and uptunneling suppress boltzmann brains. *Physical Review D*, *104*(023528), 1–8.

Page, D. N. (2024a). Bayes Keeps Boltzmann Brains at Bay. *Foundations of Physics*, *54*(62), 1–5.

Page, D. N. (2024b). Is our universe likely to decay with 20 billion years? *Physical Review D*, *78*(063535), 1–6.

Petersen, K. (1983). *Ergodic theory*. New York, NY: Cambridge University Press.

Poincaré, H. (1890). Sur le probléme des trois corps et les équations de la dynamique. *Acta Mathematica*, *13*, 1–270.

Reichert, P. (2024). The ergodic hypothesis. In A. Bassi, S. Goldstein, R. Tumulka, & N. Zanghì (Eds.), *Physics and the nature of reality* (pp. 285–299). Cham, Switzerland: Springer.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., . . . Tonry, J. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, *116*, 1009–1038.

Saad, B. (2024). Lessons from the void. *Analytic Philosophy*, *00*, 1–28.

Simone, A. D., 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*(063520), 1–30.

Stowe, K. (2007). *An introduction to thermodynamics and statistical mechanics (stowe)*. New York, NY: Cambridge University Press.

Tumulka, R. (2022). The problem of boltzmann brains and how bohmian mechanics helps solve it. In E. Battistelli, R. T. Jantzen, & R. Ruffini (Eds.), *Proceedings of the 15th marcel grossmann meeting on general relativity* (pp. 540–545). Singapore: World Scientific.

Wallace, D. (2023). A bayesian analysis of self-undermining arguments in physics. *Analysis*, *83*(2), 295–298.

Wilhelm, I. (2022). Typical. *The British Journal for the Philosophy of Science*, *73*, 561–581.

Wilhelm, I. (2025). Typicality first. *The Philosophical Quarterly*, *75*(3), 1189–1209.

Wilhelm, I. (in press-a). Typicality-based chance. *The Philosophical Quarterly*.

Wilhelm, I. (in press-b). The Typical Principle. *The British Journal for the Philosophy of Science*.

Winsberg, E. (2004). Can conditioning on the "past hypothesis" militate against the reversibility objections? *Philosophy of Science*, *71*(4), 489–504.

Zermelo, E. (1896). Über einen satz der dynamik und die mechanische wärmetheorie. *Annalen der Physik*, *57*, 485–494.