# Time's Arrow and *De Se* Probabilities

Eddy Keming Chen*

January 16, 2020

> We therefore conclude that the universe is not a fluctuation, and that the order is a memory of conditions when things started. This is not to say that we understand the logic of it.
>
> Richard Feynman (1964)

**Abstract**

One of the most difficult problems in the foundations of physics is what gives rise to the arrow of time. Since the fundamental dynamical laws of physics are (essentially) symmetric in time, the explanation for time's arrow must come from elsewhere. A promising explanation introduces a special cosmological initial condition, now called the *Past Hypothesis*: the universe started in a low-entropy state. In this paper, I argue that, in a universe where there are many copies of us (in the distant past or the distant future), the Past Hypothesis needs to be supplemented with *de se* (self-locating) probabilities. However, letting in *de se* probabilities also helps its rival—the *Fluctuation Hypothesis*, leading to a kind of empirical underdetermination and radical epistemological skepticism. The skeptical problem is exacerbated by the possibility of "Boltzmann bubbles." Hence, it seems that explaining time's arrow is more complicated than we have realized, and that its explanation may depend on how we resolve philosophical issues about *de se* probabilities. Thus, we need to carefully examine the epistemological and probabilistic principles underlying our explanation of time's arrow. The task is especially urgent for theories that invoke the Past Hypothesis. The philosophical analysis offered in the paper aims at preparing the conceptual foundation for such a task.

*Keywords: Past Hypothesis, Principle of Indifference, Principal Principle, de se probabilities, time's arrow, entropy, Bayesian epistemology, confirmation, empirical underdetermination, skepticism*

*Department of Philosophy, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093-0119. Website: www.eddykemingchen.net. Email: eddykemingchen@ucsd.edu

## Contents



# 1 Introduction

One of the most difficult problems in the foundations of physics is what gives rise to the arrow of time.[1] On the one hand, nature displays a striking pattern of temporal asymmetry. Let us think through some familiar examples. An ice cube in a cup of hot coffee will melt; gas molecules contracted to a corner of the room will spread out; and a banana on the kitchen table will turn black. These processes (and many others) have a preferred temporal direction: they only happen from the past to the future. We do not see them happen in the other direction: an ice cube spontaneously form in a cup of hot coffee; gas molecules spontaneously contract to a corner; or a banana becomes fresher after a week.

Here is another way to put it. In the past, the ice cube was larger, the gas was contracted, and the banana was fresh; in the future, the ice cube will be smaller, the gas will be spread out, and the banana will be decayed. These processes are sensitive to the past/future distinction; they are asymmetric in two temporal directions. We never see an ice cube (in a cup of hot coffee and under no outside influence) that is larger in the future and smaller in the past.

Physical states of the unmelted ice cube, the contracted gas molecules, and the

[1] In this paper, I use the phrase "the arrow of time" to designate the asymmetry of time. In particular, I mean the "entropic" asymmetry of time for which the entropy is lower in the past and higher in the future. Such asymmetries are not intended to replace the more mysterious notions such as the "flow of time" or the "passage of time." However, the latter notions may nonetheless be related to the former notions. Understanding their connection is work outside the scope of this paper.

fresh banana are less disorderly than those of the melted ice cube, the dispersed gas molecules, and the decayed banana. The former states have less *thermodynamic entropy* than the latter states. The entropic arrow of time is defined as the direction of entropy increase in time. Such an entropy increase is summarized in the Second Law of Thermodynamics:

**The Second Law** The (thermodynamic) entropy of a closed system does not decrease over time.

On the other hand, the fundamental dynamical laws of physics are symmetric in time. Take $F = ma$ for example. For any sequence of particle configurations that obeys $F = ma$, the time-reversal of that sequence also obeys $F = ma$: one simply needs to reverse the direction of the final particle velocities to get back to the initial state (with the opposite velocity). Similarly, the Schödinger equation of quantum mechanics and the Einstein field equation of general relativity are also symmetric in time (although their time-reversal operations are somewhat different). The fundamental dynamical laws of physics allow an ice cube to melt and also to spontaneously form in a cup of coffee. They are not sensitive to the past/future distinction, and they do not have a temporal directionality. Thus, they do not explain the arrow of time.

Therefore, the arrow of time cannot come from the fundamental dynamical laws alone. Indeed, the standard explanation of time's arrow makes use of a special initial condition. It is a plausible idea: if we start a physical system in a very low-entropy state (unmelted ice, contracted gas, and fresh bananas), then the dynamical laws will (almost surely) take it to a higher-entropy state.[2] However, it is rather complicated to postulate a low-entropy initial condition for every physical system. Instead, we can postulate a low-entropy initial condition for the whole universe, which (by some plausibility arguments and some rigorous mathematical proofs) will likely lead to an increase of entropy for the whole universe as well as an increase of entropy for typical subsystems in the universe. This low-entropy initial condition for the universe is now called the *Past Hypothesis* (Albert 2000).

The explanation of time's arrow in terms of the Past Hypothesis has many well-known advocates: Boltzmann (2012), Feynman et al. (2011), Lebowitz (2008), Penrose (1979), Albert (2000), Callender (2004), North (2011), Wallace (2011), and Loewer (2016).[3] Setting aside future progress in cosmology,[4] this explanation might be one of the most promising ones for understanding the time-asymmetry in our universe. Moreover, it has been argued that the explanation can also justify our standard inferences to the past.[5] Thus, the explanatory success is supposed to extend

[2]We will explain the qualifier "almost surely" when we discuss the Statistical Postulate in §2.1.

[3]See Earman (2006) for some worries about the Past Hypothesis as a hypothesis about the initial condition for the universe. See Goldstein et al. (2016) for a discussion about the possibility, and some recent examples, of explaining the arrow of time without the Past Hypothesis.

[4]See, for example, the interesting ideas of Carroll and Chen (2004). At the moment their proposal is quite speculative, but it is conceptually illuminating as a possible alternative to the Boltzmannian paradigm shared by the previously quoted people. See Goldstein et al. (2016) and Lazarovici and Reichert (2018) for some discussions of the ideas of Carroll and Chen based on simpler models.

[5]See especially Albert (2000) and Loewer (2016).

to the epistemic realm, in which our ordinary beliefs about the past obtain their justification partly in virtue of the Past Hypothesis.

However, as I shall argue in this paper, the Past Hypothesis may need to be supplemented with certain *de se* probabilities, taking the form of a self-locating probability distribution that has a temporal bias. Moreover, I argue that allowing such *de se* probabilities would open up the Pandora's box: it leads to radical epistemological skepticism.

First (in §2), I argue for the *De Se* Thesis: the Past Hypothesis explanation of time's arrow requires a postulate about *de se* probabilities. The reason for postulating certain *de se* probabilities is as follows. In a universe that persists long enough, there may be many instantaneous copies of us (e.g. by random statistical fluctuations). For reasons similar to the Boltzmann Brain argument (if we are equally likely to be any observers with our mental states, then it is most likely we are observers produced by the smallest fluctuations—the Boltzmann brains), we need to postulate that our current time is located between the Big Bang and the first equilibrium. Even though this may be familiar to experts in the foundations of statistical mechanics, I hope it will become clear that additional clarity is needed to fully understand the issues we face, given the importance and subtlety of the issues involved.

Second (in §3), I argue that we can also use *de se* probabilities to help a rival of the Past Hypothesis—the Fluctuation Hypothesis. I show that by adding suitable *de se* probabilities to the Fluctuation Hypothesis, we can make it empirically on par with the Past Hypothesis. A surprising consequence is that their empirical underdetermination could lead to radical skepticism. The augmented version of the Fluctuation Hypothesis resembles the infamous Boltzmann Brain Hypothesis. However, as I shall argue, the augmented version is more believable and epistemically robust given what I call "Boltzmann bubbles." It can even accommodate various epistemological positions that are ruled out by the Boltzmann Brain Hypothesis.

The skeptical conclusion puts pressure on the view that we should add *de se* probabilities to explain the arrow of time. This is compatible with the possibility that the skeptical conclusion might be avoided by appealing to certain traditional epistemological responses. Thus, I hope the present analysis will be interesting to epistemologists and general philosophers of science.

In this paper, I assume that the universe allows many thermodynamic fluctuations. It is an open question whether there are such fluctuations in our universe, depending on whether the universal state space is finite or infinite (in terms of phase space volume or Hilbert space dimension). But even if this were not true of the actual universe, I hope the reader will be convinced that it would still be worth investigating how serious the problem would be, and what strategies we would need, if Nature were not so kind to us.

## 2 The *De Se* Thesis

In this section, we will first review the standard explanation for time's arrow in terms of the Past Hypothesis. We will introduce some concepts from philosophy of physics that may be unfamiliar to non-specialists. We will then consider whether it is the best theory that explains the evidence. To do so we will consider an alternative explanation without the Past Hypothesis. Thinking about their differences leads us to the *De Se* thesis. Some of these discussions will be familiar to people working on the foundations of statistical mechanics. Not enough has been said in the philosophical literature in a way that connects to principles of probabilistic inferences. By making explicit the underlying assumptions and philosophical principles, I hope to provide a new philosophical analysis of the epistemic situations surrounding the Past Hypothesis, thermodynamic fluctuations, and Boltzmann brains. Given the complexity and subtlety of the issues, we ought to strive for clarity and honesty so that we do not get ourselves in unnecessary confusions or gloss over important distinctions. Once we have a clear understanding of the problem and the assumptions leading to it, we might come to see what kind of solutions are needed.

### 2.1 The Mentaculus Theory

According to the standard picture, the Past Hypothesis is key to understand the apparent temporal asymmetry: from ice melting and gas dispersing to the more general statement about entropy increase in the Second Law of Thermodynamics. However, it is to be supplemented by two more postulates: the dynamical laws of physics (such as $F = ma$) and a probabilistic postulate called the *Statistical Postulate*. Together, they provide the standard (probabilistic) explanation of time's arrow, i.e., the unidirectional change of entropy.

To appreciate the explanation that the Past Hypothesis provides, let us introduce some technical terms from statistical mechanics. The Past Hypothesis ensures that the universe started in a low-entropy initial condition.[6] But how should we understand this notion of entropy? Entropy is a macroscopic quantity that is on par with density and pressure. It can be calculated by measuring the transformations in thermodynamic quantities.[7] However, it can also be defined using a distinction between the macrostates and the microstates. To illustrate, let us consider a classical-mechanical system of gas molecules in a box.[8] A macrostate is a characterization

[6]In this paper we assume that the universe in fact has a temporal boundary or a space-time singularity that can be called "the beginning." To relax this assumption, as some cosmological theories does, will take us to more complex issues than we have the space to discuss here. The lesson we learn under such an assumption might carry over (with some modifications) to the more general discussion that does not assume that.

[7]For example, Clausius (1867) defines the change in entropy for an isolated system to be equal to the change in heat divided by temperature.

[8]Below we will use some concepts from classical statistical mechanics. For the technically inclined, we note that the classical framework can be adapted to a quantum framework as follows: use a Hilbert space instead of phase space, a state vector instead of a phase point, a subspace instead of a subset, and dimension counting instead of volume measure to measure entropy. See Goldstein and

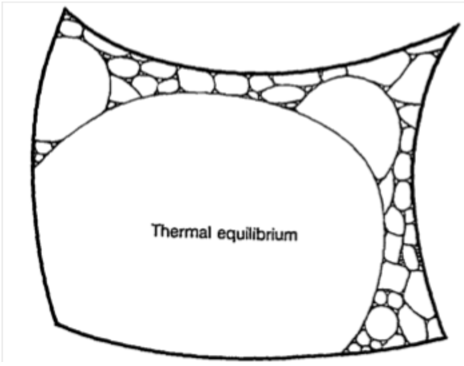


Figure 1: The phase space (restricted to the energy shell). Each point corresponds to an exact microstate. Each bounded region corresponds to a macrostate, which is a set of macroscopically indistinguishable microstates. The equilibrium macrostate (maximum entropy) takes up overwhelmingly most volume; all other macrostates are much smaller. The macrostates partition the phase space into regions. From Penrose (1989).

of the gas in a box in terms of macroscopic variables such as pressure, volume, and temperature, while a microstate is a characterization of the system in terms of microscopic variables such as the positions and velocities of all of the gas molecules. A macrostate is compatible with many possible microstates—a macrostate is multiply-realizable by many different microstates. The actual microstate realizes a particular macrostate, but it shares the macrostate with many other microstates (that are macroscopically similar).

Intuitively, there are more microstates realizing the macrostate in which all the gas molecules are spread out uniformly than there are microstates realizing the macrostate in which all of them are contracted to a corner. According to Ludwig Boltzmann, (thermodynamic) entropy measures how "many" microstates are compatible with a given macrostate.[9] It follows, on this definition of entropy, the "spread-out" state has higher entropy than the "contracted" state. The uniformly spread-out state of gas in a box also corresponds to thermal equilibrium, the state of entropy maximum. We can plot all the microstates on a $6N-$dimensional space called the *phase space*. The macrostates are sets of microstates that are macroscopically indistinguishable; they partition the space into distinct non-overlapping regions. The standard way of

Tumulka (2011) for an informative overview of Boltzmannian quantum statistical mechanics.

[9] For the technically inclined, here are some formal details. The Boltzmann entropy of a microstate is proportional to the volume of the macrostate that it belongs to:

$$S_B(X) = k_B \log|\Gamma_{M(X)}|,$$

where $X$ is the microstate of the system, $k_B$ is the Boltzmann constant, $\Gamma_{M(X)}$ is region of phase space that corresponds to the macrostate of $X$, and $|\cdot|$ denotes the volume measure of the $6N$-dimensional phase space. It follows that the larger the macrostate, the higher the Boltzmann entropy. By Liouville's Theorem, the Lebesgue measure is invariant under the classical equations of motion. If at $t_1$ the system is in $X_1$ which belongs to a small macrostate $M_1$, and at $t_2$ the system is in $X_2$ which belongs to a large macrostate $M_2$, then the Boltzmann entropy has increased from $t_1$ to $t_2$. The transition from $X_1$ to $X_2$ from $t_1$ to $t_2$ is determined by the classical laws of motion, i.e. Hamilton's equations.

counting the microstates, since there is an infinity of them, is by using the standard Lebesgue measure on phase space. On this way of counting, the equilibrium state takes up the overwhelming majority of volume in phase space (see Figure 1).

In the language of phase space, we can consider (as a first approximation) the entire universe to be such a classical system. Then the Past Hypothesis selects a special macrostate $M(0)$ to be where the initial microstate of the universe lies in. $M(0)$ is small in volume measure. Thus, it has very low entropy (by Boltzmann's definition). But it is still compatible with a continuous infinity of microstates. For a typical initial microstate lying in $M(0)$, it will follow the dynamical laws (such as $F = ma$) to evolve into other microstates. Since the overwhelming majority of microstates surrounding $M(0)$ will lie in macrostates of larger volume, typical trajectories from $M(0)$ will get into larger macrostates than $M(0)$, which correspond to higher entropy.[10]

However, the Past Hypothesis is still not sufficient. Not all microstates lying in $M(0)$ will get into higher-entropy macrostates. Some of them are "bad": they will evolve under the dynamical laws into lower-entropy macrostates.[11] We need a reason to neglect these microstates. The Statistical Postulate provides such a reason. It specifies a uniform probability distribution (with respect to Lebesgue measure) on phase space. According to this probability distribution, these "bad" microstates are overwhelmingly unlikely. That is, it is overwhelmingly likely that our world did not start in one of those "bad" microstates. Hence, it is with overwhelming likelihood that the entropy of our world has always been increasing in the past and will continue increasing in the future. (See Figure 2.) This constitutes a probabilistic explanation of the entropic arrow of time.

Following Albert (2000) and Loewer (2016), let us call the following the Mentaculus Theory ($T_M$)[12]:

1. **Fundamental Dynamical Laws (FDL):** A specification of the evolution of the fundamental microstates of the universe (and the fundamental microstates of its isolated sub-systems).

2. **The Past Hypothesis (PH)**: A specification of a boundary condition characterizing the universe's macrostate at the time of the Big Bang as M(0). In agreement with contemporary cosmology, M(0) is a macrostate with extremely low entropy.

[10] This is a place where rigorous results are difficult to obtain. The history of statistical mechanics contains many attempts to make progress in this direction. Boltzmann's not fully rigorous argument for his Boltzmann equation is one step in the direction. Oscar E. Lanford's celebrated proof of statistical results in a model of hard spheres and diluted gas are further steps of significance. See Uffink and Valente (2010) for more discussions and references.

[11] Their existence is suggested by the time-reversal invariance of the dynamical laws.

[12] The term "Mentaculus Theory" comes from the phrase "Mentaculus Vision" which is coined by Loewer (2016). It is based on the movie *A Serious Man* (2009) directed by Ethan Coen and Joel Coen. In the movie, the title of Arthur's book is *The Mentaculus*, which means the "probability map of the universe." Here, the Mentaculus Vision is supposed to provide, given the probability distribution of PH + SP, a probability assignment of every proposition that can be formulated in the languages of phase space or of the Hilbert space. Barry Loewer calls the joint system—the package of laws that includes PH and SP in addition to the dynamical laws of physics—the Mentaculus Vision.

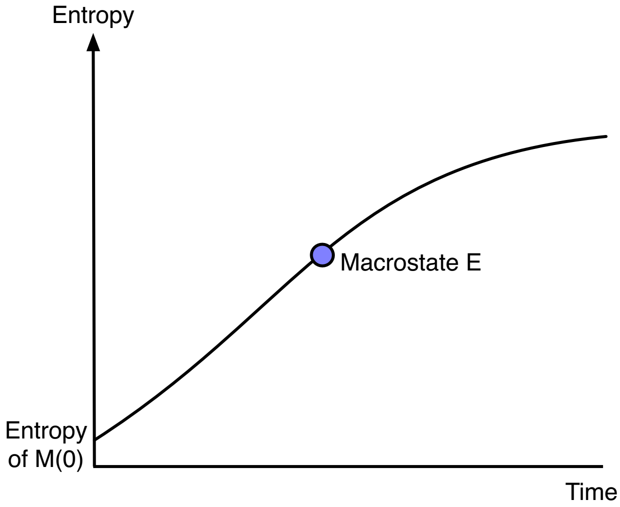


Figure 2: The Past Hypothesis selects a low-entropy initial condition, a special macrostate $M(0)$. The macrostate $E$ stands for the macrostate of the universe at the time of observation.

3. **The Statistical Postulate (SP)**: A uniform probability distribution (specified by the standard Lebesgue measure) over the physically possible microstates that realize M(0).

Together, these three postulates provide an explanation for the arrow of time exemplified by the unidirectional change of entropy. PH and SP constitute an extremely biased initial probability distribution on phase space. They make it overwhelmingly likely that the microstate of our universe lies on a trajectory that will go to higher-entropy states.

## 2.2 The Fluctuation Theory

The Mentaculus Theory ($T_M$) explains the origin of temporal asymmetry in terms of a specially chosen initial condition and a probability distribution—the Past Hypothesis (PH) and the Statistical Postulate (SP). It would have been a powerful explanation had it been successful, for it would have explained a large class of phenomena: ice melts in room temperature; things grow and decay; and many other temporally asymmetric phenomena (and perhaps even the asymmetry of records and control: that we have records of the past but not of the future, and we currently have control of the future but not of the past[13]). This theory could be our best guide to time's arrow.

What are our grounds for endorsing $T_M$? Defenders of this theory (such as Albert and Loewer) suggest that it is our best explanation for time's arrow. But can we really

[13]See Albert (2000) and Loewer (2016) for arguments that connect the Mentaculus Theory to these other arrows of time.

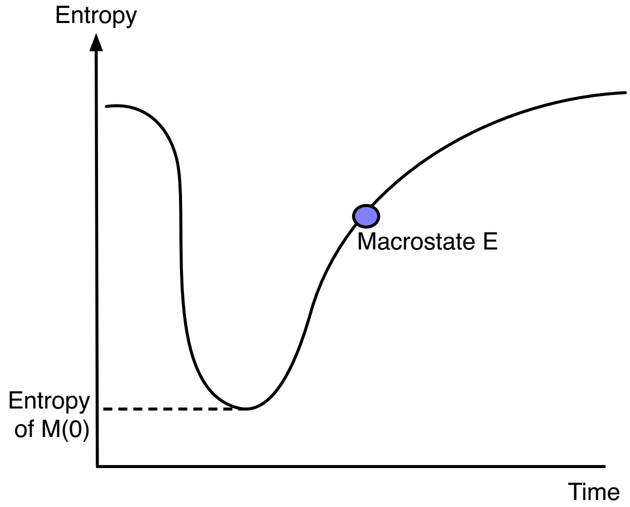


Figure 3: The Fluctuation Hypothesis allows the universe to start in any macrostate. But the theory predicts that the universe will eventually fluctuate to a low-entropy macrostate $M(0)$, which produces the history of the observable universe, including the macrostate $E$ at the time of observation.

rule out all other possible explanations? Is PH really better than its alternatives? The answers are not so simple.

Let us consider an alternative to $T_M$—the Fluctuation Hypothesis. This alternative also goes back to Ludwig Boltzmann (2012 (1899), §90). He might even have endorsed it himself at some point. The motivation is as follows. It seems that the special initial condition in $T_M$ is too special and too contrived. Penrose (1989) estimates that the initial macrostate $M(0)$ is tiny compared to the available volume on phase space. A calculation based on classical general relativity suggests that the Past Hypothesis macrostate (specified using the Weyl curvature) is only $\frac{1}{10^{10^{123}}}$ of the total volume in phase space. It seems more satisfactory to explain why the universe started in such a special state than to postulate it as axiomatic as on $T_M$.

One way to explain PH is based on thermodynamic fluctuations. The universe, considered as a closed thermodynamic system, will typically increase in entropy and remain the same after it reaches the entropy maximum. However, sometimes it will decrease in entropy, producing a thermodynamic fluctuation—a deviation from the normal behavior. Thermodynamic fluctuations are rare, but they do occur given enough time. This is the reason that the Second Law of Thermodynamics should be regarded as a probabilistic (or statistical) law that holds with overwhelming probability but not with certainty.

An extreme kind of fluctuation is demonstrated by Poincaré to occur in some systems. The Poincaré Recurrence Theorem says, roughly, that if we start from anywhere in phase space, we will (almost surely) come back to it infinitely many times.[14] The argument, however, assumes that the phase space of the system is bounded with finite volume. This assumption may or may not apply to our universe. It is still an open question whether there can be such dramatic kinds of fluctuations, i.e. recurrences, in our universe. The question depends on whether the universal state space (phase space or Hilbert space) is infinite. However, even if we cannot guarantee the existence of recurrences by something like Poincaré Recurrence Theorem, we cannot rule out the existence of less dramatic fluctuations and more localized fluctuations. Recurrences are sufficient but not necessary for the Fluctuation Hypothesis, which only attempts to explain our observation by some random fluctuation (which can be much smaller than a full recurrence). In any case, this is not the place to settle the technical question.[15] The goal of the paper is rather conceptual. As mentioned earlier, even if we were to have empirical grounds to entirely rule out the possibility of fluctuations in the actual universe, it would still be interesting to investigate how serious the problem would be if fluctuations were possible.

Assuming that there are suitable fluctuations in the universe, then we can use them to explain the origin of the initial low-entropy state described by the Past Hypothesis. The universe started in some generic microstate $x_0$ "chosen at random" from phase space (restricted to the energy hypersurface). Most likely it started in thermal equilibrium, and it will stay in that state for a long time. However, given enough time, it will fluctuate into lower-entropy states. Eventually, it will fluctuate into an extremely low-entropy state—the macrostate $M(0)$ selected by the Past Hypothesis. From that state, the universe will grow in entropy, as we have explained on $T_M$. (See Figure 3.) Therefore, the Fluctuation Hypothesis can also explain time's arrow, and it does so without postulating a special initial condition. To summarize, the Fluctuation Theory ($T_F$) consists in the following postulates:

1. The fundamental dynamical laws.

2. A uniform probability distribution over all microstates (on the energy shell).

Hence, $T_F$ is essentially $T_M$ without PH. To see this, we note that SP can be understood as the uniform probability distribution that is conditionalized on any initial condition. In $T_M$, SP is conditionalized on PH. In $T_F$, there is no special initial condition, so the probability distribution remains completely uniform over all microstates on phase space.

[14]For the mathematically inclined, here is a rigorous statement of the theorem. Let $(X, \mathscr{B}, \mu)$ be a measure space. Let $T : (X, \mathscr{B}) \to (X, \mathscr{B})$ be a function such that $\forall A \in \mathscr{B}, T^{-1}(A) \in \mathscr{B}$. Definition: the measure $\mu$ is a $T$-invariant measure if $\forall A \in \mathscr{B}, \mu(A) = \mu(T^{-1}(A))$.

**Theorem 2.1** (The Poincaré Recurrence Theorem) *Let $\mu$ be a $T$-invariant measure with $\mu(X) < \infty$. $\forall A \in \mathscr{B}$ such that $\mu(A) > 0$, we have a.e. $x \in A$ such that $\#\{n \in \mathbb{Z}^+ | \; T^n(x) \in A\} = \infty$.*

[15]For some recent work, see Carroll (2017) for discussions about the possibility of a finite-dimensional Hilbert space.

## 2.3 *De Se* Probabilities

How should we adjudicate between $T_M$ and $T_F$, when both seem to explain time's arrow? One could appeal to super-empirical virtues. For example, $T_F$ seems much simpler than $T_M$: it has fewer axioms. Moreover, $T_F$ seems less *ad hoc* than $T_M$: in explaining the temporal asymmetry, it does not break temporal symmetry by adding a special initial condition (PH).

However, things are much more subtle than they seem. There are two lines of reasoning that are often considered. On the first line of reasoning, it seems that $T_F$ is much worse than $T_M$. It is rare to have thermodynamic fluctuations, and it is extremely rare to have fluctuations that produce an extremely low-entropy state such as $M(0)$. Intuitively, therefore, on $T_F$ it is extremely unlikely to find ourselves living in the current state—a medium-entropy state not too long after the Big Bang and some time away from thermodynamic equilibrium. Most likely, the intuition goes, we would find ourselves in the equilibrium, which is contrary to our evidence.

On the second line of reasoning, $T_F$ is no worse than $T_M$. Although it is true that large fluctuations are infrequent, they do occur given enough time. Indeed, large fluctuations will occur with probability close to 1 if we wait long enough. Therefore, in a universe described by $T_F$, some creatures will find themselves in a state that is exactly like our current macrostate.

Both lines of reasoning seem plausible. The difference is that they are tracking two kinds of probabilities: probability *de dicto* and probability *de se*. To appreciate this distinction, we need to understand the distinction between proposition *de dicto* and proposition *de se*. A proposition *de dicto* describes the world in objective terms; an example is "planet Earth is in the solar system." A proposition *de se* involves indexical terms; examples include "I am on planet Earth right now." Familiar examples from Perry (1979) and Lewis (1979) suggest that one can have knowledge of a *de dicto* proposition without having knowledge of the corresponding *de se* proposition. For example, Katy who has amnesia can know that Katy is in Paris without knowing that she herself is in Paris. Probability *de dicto* is a probability function over *de dicto* propositions, with the probability space being the space of possible worlds. Probability *de se* is a probability function over *de se* propositions, with the probability space being the space of "centered worlds" (Lewis (1979)), which are possible worlds with "centers" being certain agents or perspectives. For example, we can imagine a centered world with two qualitatively identical copies of Katy, one on planet Earth and the other on planet Twin Earth, but the "center" is on planet Earth. That could represent (from Katy's perspective) the *de se* proposition "I am on planet Earth." Probability *de se* is sometimes also called *self-locating* probability, as it can be used to describe my uncertainty over where I am in space or time.

The second line of reasoning tracks probability *de dicto*: the probability that there is a fluctuation in the history of the universe producing the $M(0)$ state is close to 1. In contrast, the first line of reasoning tracks probability *de se*: the probability that we find ourselves in such a fluctuation is almost zero, given that fluctuations are so rare and the history of the $T_F$ universe is almost entirely in thermodynamic equilibrium.

But the latter judgment must implicitly rely on some sort of principle of indifference over *de se* propositions: the probability that we are in any particular time-interval is proportional to the length of that interval. For example, the (unconditional) probability that we find ourselves in the first billion years of the universe's history is the same as the probability that we find ourselves in the second billion years of the universe's history, and so on. Since the overwhelming majority of times is taken up by thermodynamic equilibrium and not by fluctuations, it is extremely unlikely that we find ourselves in a fluctuation.[16]

Since a lot depends on what we mean by "our current evidence," we should clarify what our current evidence consists in. And by that I mean our *direct* evidence, as we are attempting to examine the justification for our inferential beliefs, including the two hypotheses. Since we are in the context of scientific reasoning, we should be neither too stringent nor too permissive. It seems that we should be realists about the external world. That is, we are justified in believing that we are not BIVs or Boltzmann Brains. However, it seems inappropriate to take ourselves to have direct access to the exact microstate of our current universe, any states of the past universe, or any states of the future universe. All of them are usually inferred from our direct evidence. For our discussions below, we can entertain different ideas about our evidence that satisfy those constraints. But for concreteness, as a first approximation, we stipulate that our current evidence consists in what Albert (2000) calls the *directly surveyable* condition of the world currently happens to be, i.e. the macrocondition of the world at this instant, which includes, for example, the locations and configurations of galaxies, planets, tables, chairs, observers, pointers used in detection devices, photographs, and newspapers.[17] The central question is whether such evidence is sufficient epistemic ground for either hypothesis. Let us take our current evidence $E$ to be the following:

**Evidence:** Our current evidence $E$ is the medium-entropy macrostate of the universe at this moment.

How do $T_M$ and $T_F$ compare with respect to our current evidence $E$? Suppose we use the Bayesian framework to compare theories. Which one has higher *posterior* probability? According to a simple argument below, it seems that $T_M$ wins the competition:

---

[16]This discussion somewhat resembles the debate about fine-tuning argument for design and the multiverse response. The multiverse proponent invokes the *anthropic principle*, emphasizing on the observation-selection effect. The "this universe" objection to the multiverse response focuses on the *de se* element in our evidence, which seems to track a different kind of probability. See, for example, White (2000) and Manson and Thrush (2003) for discussions.

[17]See Albert 2000, p.96. This is obviously too generous. But it will simplify things. The arguments below are robust with respect to reasonable relaxation of this condition.

**The Master Argument**

P1 Our current evidence $E$ is much less likely on $T_F$ than on $T_M$:

$$P(E|T_F) \ll P(E|T_M).$$

P2 $T_F$ is roughly as intrinsically likely as $T_M$:

$$P(T_F) \approx P(T_M).$$

C $T_F$ is much less likely than $T_M$ given our current evidence $E$:

$$P(T_F|E) \ll P(T_M|E).$$

The Master Argument is valid by an application of Bayes's theorem (assuming $P(E) \neq 0$):

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

For now, we have adopted a "uniform" prior probability over temporally *de se* propositions: we could be anywhere in time, with a probability distribution that is uniform over the entire history (or more or less flat).[18] Let us call this assumption PoI-De-Se (inspired by a restricted version of the principle of indifference). This assumption is simple, but it is arguably an oversimplification. If we do not exist, we cannot observe the universe. Any agent considering what kind of prior probability she should adopt should also take into the *a priori* fact that she does exist and have certain conscious experiences. Suppose materialism about the mind is true, then her existence and experiences would require the existence of certain physical states. Suppose a thinking brain (or something like it) is the minimal requirement for her existence and experience. By taking into account her own existence and experience (which is knowledge *a priori*), she ought to rule out those time intervals containing no brains or no brains that have her experiences. This will produce a biased probability distribution—she is equally likely to be any brain with her conscious experiences as any other brain. Call this distribution PoI-De-Se*.

Assuming PoI-De-Se*, then, the first premise of the Master Argument does not go through so easily. We are not as likely to be in any interval in time as any other interval. Our temporal location is restricted to those intervals where there are conscious beings with our experiences, which is a severe restriction. However, another problem arises. The vast majority of fluctuations in the universe are quite

[18]Technically speaking, for a universe without temporal boundaries, the uniform distribution is no longer normalizable. To have a normalizable probability distribution, we could use a Gaussian distribution that is more or less flat, centered on some point in time. This introduces a temporal bias near the center of the Gaussian. This problem of non-normalizability comes up frequently in cosmology when we consider infinite models. Ideally we would like to find a particular natural choice of measure or probability distribution, which may not exist. For an interesting philosophical discussion on non-normalizable measures, responses, and possible pitfalls, see McGrew et al. (2001) and the references therein.

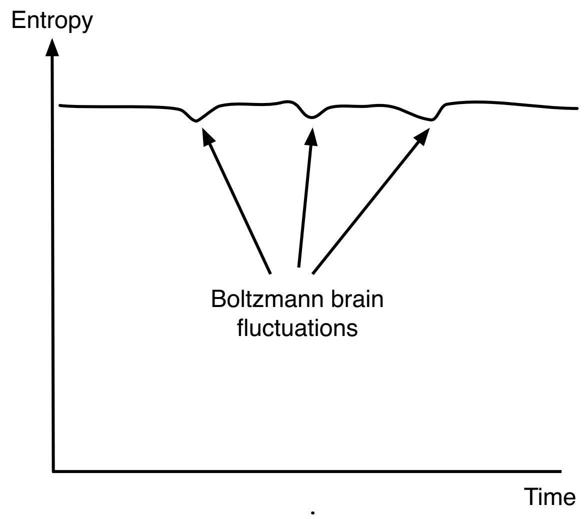


Figure 4: Boltzmann Brains

small. Typically, they are small deviations from thermal equilibrium. It is very rare for the universe to fluctuate into something like a Big Bang state. It is exponentially more common for it to fluctuate just into the current macrostate. Moreover, it is exponentially more common for it to fluctuate just into a state with no structure at all except for a number of brains thinking the same thoughts we do now (and even more common for it to fluctuate into a state with just one brain). (See Figure 4.) They would have all the memories of the past (which would be false) and all the perceptions of the present (which would be non-veridical). These unfortunate beings are called *Boltzmann brains*. Since there are exponentially more Boltzmann brains than ordinary observers, by PoI-De-Se*, we should be much more confident that we are Boltzmann brains, which is absurd. Let us call this *the Boltzmann brains problem*.

In our context, the Boltzmann brains problem comes up as we try to fix $T_F$ by adding a biased *de se* probability distribution. That $T_F$ has this problem seems to be another argument for $T_M$. But is it? Unfortunately, things are not so neat. $T_M$ also faces its own version of the Boltzmann brain problem.

For a universe starting from a low-entropy state (given the Past Hypothesis), it will with overwhelming probability (given the Statistical Postulate) increase in entropy. However, after it reaches thermodynamic equilibrium, the highest level of entropy, it will fluctuate downward in entropy if we wait long enough. The smallest such fluctuations compatible with our conscious experiences, which are also the most frequent, will again be the Boltzmann brain fluctuations. They will be the minimal fluctuations compatible with our experiences, ones in which there are a few brains floating temporarily in an environment that is otherwise devoid of any structure. Given $T_M$, it is nonetheless the case that typical observers (with our experiences) in the universe will be Boltzmann brains. If PoI-De-Se* is the correct *de se* probability distribution, we are most likely Boltzmann brains, which is absurd.

Here is the upshot of our discussion so far: both $T_M$ and $T_F$ face the Boltzmann brains problem. It is really a problem about temporal self-location. So we will also call it *the de se problem*. To be sure, the problem may seem to be just another skeptical hypothesis. But it is worth thinking about what one's response ought to be in this particular case. To try another response: suppose we defend $T_M$ by stipulating that we are not Boltzmann brains, and that we are ordinary observers living in the actual macrostate $E$, which contains not just a few brains but also the normal kind of environment—planet Earth, the solar system, the Milky way, etc. Suppose further that we have a uniform self-locating probability distribution over the occurrences of macrostate $E$ (which we assume will occur many times in the long history of the universe). Call this distribution PoI-De-Se**. This is still not sufficient for getting the monotonic increase of entropy for $T_M$. The minimal fluctuations compatible with $E$ are the medium-entropy dip from thermodynamic equilibrium, in which $E$ is the local minimum of entropy and the entropy is higher in both directions of time. Again, there will be overwhelmingly more minimal fluctuations than large deviations that first produce a low-entropy state described by the Past Hypothesis and then increase in entropy all the way to $E$. Hence, by the lights of $T_M$ and PoI-De-Se**, it is most likely that we have symmetric histories, just as on $T_F$ and PoI-De-Se**.

How, then, should defenders of $T_M$ get out of the present conundrum and predict a monotonic increase of entropy? It seems that the best strategy is to entirely abandon these versions of PoI-De-Se. Instead, they can choose a much more biased distribution:

**NPH:** We are currently located in the first epoch of the universe—between the time when the Past Hypothesis applies and the first thermodynamic equilibrium; we are equally likely to be any ordinary observers (inside the first epoch) that have our current experiences.

The last clause of uniform distribution is to make it less specific. As usual, there is a trade-off between informativeness and simplicity. Here we have already made the distribution extremely specific—about the first epoch. However, in order to have a good theory, we should make the postulate as generic as possible given that constraint. Thus, we postulate a uniform *de se* distribution on this interval.[19]

Given NPH, our temporal location is restricted to the first epoch of the universe, between the time when the Past Hypothesis applies and the first equilibrium. Thus, we are in the period of "normal history," where fluctuations are probably non-existent. Given our current evidence $E$, then we can predict that our past had lower entropy and our future will have higher entropy, which meets our goal to predict a monotonic increase of entropy around us. NPH is very informative—perhaps too informative for a self-locating distribution. It is intended as an objective norm for the self-locating probability that goes beyond simple indifference.[20] It would be even more compelling

[19] Winsberg (2012) and Loewer (2016) postulate a version of NPH for which we are located in the first epoch of the universe. It is essentially the same idea. However, they do not explicitly connect them to *de se* probabilities.

[20] There is another place we could postulate such norms—a Many-Worlds interpretation of quantum

if we can derive it from more self-evident principles about rationality. But it is hard to see how such a derivation would go. Absent such derivations, we should understand NPH as a fundamental postulate to go with $T_M$.

To be sure, it seems odd to have a postulate like NPH in the fundamental theory of physics, even though it plays the same role as explaining temporal asymmetry as the Past Hypothesis and Statistical Postulate. But NPH seems to be exactly what is needed to save $T_M$. In any case, notwithstanding its *de se* character, NPH gets the work done for $T_M$.[21] Given NPH and our evidence $E$, $T_M$ predicts the correct temporal asymmetry of monotonic entropy growth. Let us use $T_M^*$ to designate the combined theory of $T_M + NPH$. Hence, we have arrived at the *De Se* Thesis:

***De Se* Thesis** The explanation of time's arrow by $T_M$ requires an additional postulate about *de se* probabilities.

## 3 Radical Skepticism

In this section, we suggest that, if it is permissible to save $T_M$ with *de se* probabilities, we may open a new door to radical epistemological skepticism. First, I will consider a strategy for resurrecting the Fluctuation Theory by using self-locating probabilities. Second, I will introduce "Boltzmann bubbles." Third, I will present a new version of the Master Argument according to which the revised Fluctuation Theory is on a par with, if not better than, the revised Mentaculus Theory $T_M^*$. Fourth, I will argue that this leads to radical skepticism. Finally, I will consider some possible responses to the skeptical conclusion.

### 3.1 The Medium Entropy Hypothesis

We saved $T_M$ from the *de se* problem by choosing a temporally biased self-locating distribution—NPH. We will now argue from parity and show that we can choose another self-locating distribution to save $T_F$ from the *de se* problem.

Recall that $T_F$ has lower posterior probability given $E$ than $T_M$, because $T_F$ assigns low probability to the *de se* proposition $E$ assuming some plausible versions of the principle of indifference. If we can find a way to make $E$ as probable on $T_F$ as on $T_M^*$, then $T_F$ would be on a par with $T_M^*$ with respect to our evidence. In fact, a natural strategy exists. Let us add to $T_F$ the following Medium Entropy Hypothesis:

mechanics in which the Born rule is a self-locating probability not derived from simple axioms but put in by hand. This might be the best strategy forward if none of the existing derivations is compelling. See Sebens and Carroll (2016) for an interesting recent attempt of deriving the Born rule from other epistemic principles.

[21] The *de se* character of NPH raises many questions. Since it is unlikely to be derivable from anything else, should we treat it as a fundamental axiom in science as a fundamental law? Alternatively, should we treat it as merely a rationality principle? We will leave it for future work how to think about the nature of these *de se* probabilities.

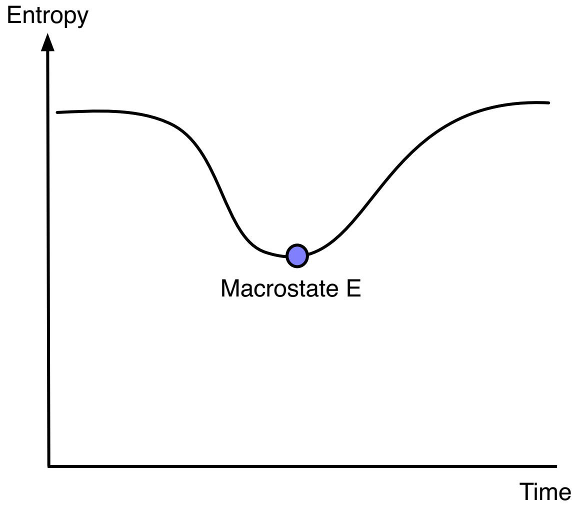


Figure 5: Medium Entropy Fluctuations

**MEH:** We are currently located in a medium fluctuation of a special kind; we are in a fluctuated state of medium entropy and strong correlations; we are equally likely to be any observers (inside these states) that have our current experiences.

MEH is similar to NPH, the self-locating postulate in $T_M^*$. Without MEH, and with only some versions of PoI-De-Se, $T_M$ predicts that we are most likely Boltzmann brains, and we are the results of tiny fluctuations. With MEH, our temporal location is restricted to temporal intervals with medium-entropy fluctuations, ones that are much larger deviations from equilibrium than Boltzmann brain fluctuations (see Figure 5). However, they are not as large as the extremely large deviations required to produce a low-entropy state described by the Past Hypothesis. Even so, given that we are located in certain medium-entropy fluctuations, we are guaranteed that we are most likely not Boltzmann brains.[22]

(One might reasonably ask: Is MEH the most natural strategy? Why not just postulate that we are currently located in a period of relaxation following a large fluctuation that resembled the initial state described by the Past Hypothesis? The short answer is that such a large fluctuation is overwhelmingly less frequent to occur than medium fluctuations that just produce the current macrostate. Among versions of the Fluctuation Hypothesis compatible with our current evidence being $E$, other things being equal, our prior distribution should favor those versions that postulate we are more likely to be located in shorter and smaller fluctuations than those that postulate we are more likely to be in longer and larger fluctuations. We will come back to this point in §3.2 and §3.5.)

[22]Given enough time, there will be some medium-entropy fluctuations that are like $E$ but also contain small local fluctuations of Boltzmann brains with our conscious experiences. But these cases are extremely rare. Given the "uniform" probability distribution over observers with our conscious experiences, we are very unlikely to be Boltzmann brains.

The requirement that the medium fluctuations have to display strong correlations is to make the theory predictively equivalent to the Past Hypothesis. There are many medium-entropy fluctuations that are abnormal from our point of view, i.e. there are no correlations among different parts of space.[23] For example, a medium-entropy fluctuation may contain a photograph of Barack Obama but no real person of Barack Obama; it may contain a left shoe of Napoleon but no right shoe of Napoleon; it may contain a book about pyramids but no real pyramids. These medium-entropy fluctuations, although devoid of the usual correlations among things in space, can nonetheless have medium level of entropy. Their amount of disorder is not lower or higher than the present macrostate $E$. But they are dramatically different from $E$. In fact, most medium-entropy fluctuations are without the right kind of correlations we are used to. That is why we need to add the condition that we are located in the right kind of fluctuations—ones that not only have the right level of entropy but also display strong correlations.

However, it may be complicated to specify exactly what the correlations are. In $T_M^*$, the Past Hypothesis plays an important role in explaining the correlations. Possible microstates coming out of the Past Hypothesis macrostate will become worlds with the right kind of correlations—they will start from a low-entropy state of hot, dense, contracted cosmic soup and will evolve into states with galaxies and more structures. The correlations are already built into the selection of a special low-entropy initial macrostate. We can exploit the Past Hypothesis to define the correlations in the following way:

**Strong Correlations:** The relevant medium-entropy fluctuations are those that produce macrostates that display the same kind of correlations[24] as if they evolved from the Past Hypothesis initial condition in the first epoch of the universe.

That is, the macrostates produced by medium-entropy fluctuations are exactly those macrostates allowed by NPH. Let us call macrostates allowed by NPH *normal macrostates*. These are the kind of macrostates with strong correlations. Thus, we can change MEH into the following:

**MEH′:** We are currently located in a *normal macrostate* produced by a medium fluctuation; we are equally likely to be any observers (inside such macrostates) that have our current experiences.

By locating ourselves in medium fluctuations that produce *normal macrostates*, MEH′ provides a probabilistic boost to the *de se* proposition that $E$ is our current evidence. In fact, since the possible macrostates are exactly the same on the two theories, the probability that $T_F + MEH'$ assigns to $E$ is exactly the same as that assigned by $T_M^*$. Let us use $T_F^*$ to designate the combined theory of $T_F + MEH'$.

[23] Thanks to Tim Maudlin for pointing this out to me.

[24] One might object that the "same kind of correlatioins" is vague. However, any admissible changes to the meaning will not make much of a difference. It is worth remembering that the Past Hypothesis as formulated in Boltzmannian statistical mechanics is also a vague postulate. On pain of imposing too much sharpness in nature, the partition of phase space into macrostates (and the selection of a particular low-entropy initial macrostate), as introduced in §2.1 and Figure 1, is a vague matter.

## 3.2 Boltzmann Bubbles

Before revisiting the Master Argument, let us pause and think where we are and compare the current situation with that of the Boltzmann brains. By the lights of $T_F^*$, most likely we are in a state of "local entropy minimum," for which entropy is higher in both directions of time. It is produced by a medium-entropy fluctuation of the right sort described by MEH'. As explained before, the fluctuation is not the same kind that produces Boltzmann brains. The current macrostate has the right sort of structure as we would believe—football stadiums, motorcycles, Jupiter, the Milky way, and etc. Most importantly, there are human beings with physical bodies attached to their brains thinking and living in normal kinds of environment. The current macrostate is a fluctuation, but it is much larger than the Boltzmann brain fluctuations. Thus, the fluctuation has the right sort of structure extended in space. We will call it a *Boltzmann bubble fluctuation.* Given just our present evidence $E$, the Boltzmann bubble is an instantaneous macrostate which describes a normal spatial configuration. The minimal fluctuation compatible with a Boltzmann bubble is going to be one in which the Boltzmann bubble is the local entropy minimum, and entropy is higher in both directions of time.

In contrast to the Boltzmann brain scenarios, the introduction of Boltzmann bubbles makes $T_F^*$ much more epistemically robust. Suppose we have strong philosophical reasons to believe that we are not brains in vats. Then we may have similar strong reasons to believe that we are not Boltzmann brains. (The analogy is not perfect. Unlike BIVs, the existence of Boltzmann brains in the actual world is a salient feature given current physics.) However, our philosophical intuitions against the possibility that we are in an instantaneous Boltzmann bubble that is extended in space and compatible with our current evidence $E$ are likely weaker and less clear-cut. Even though it is very implausible that we are Boltzmann brains, it may be less implausible that we are in a Boltzmann bubble. If one is worried the temporal duration of Boltzmann bubbles, we can add that Boltzmann bubbles do not need to be short-lived (unlike typical Boltzmann brains); they can be extended in time. That is, a medium-level fluctuation may produce a state that has lower entropy than $E$ but develops into $E$ after five days. Can we be in a Boltzmann bubble that is macroscopically indistinguishable from one coming out of $T_M^*$ but only has five days of "normal" history? Perhaps we would want a longer "normal" history. What about a Boltzmann bubble that has five years of "normal" history? By this line of thought, it soon becomes unclear where we draw the line. It is unclear that we can *a priori* rule out the possibility (or to assign very low credence) to a Boltzmann bubble with an extended period of normal history. Hence, it is unclear that we can *a priori* rule out the possibility that we are in an Boltzmann bubbles.

One might prefer to live in a Boltzmann bubble that stretches all the way back to a low-entropy state (such as the same initial macrostate described by the Past Hypothesis) such that the current macrostate is about 14 billion years away from the entropy minimum of that fluctuation. To live in such a bubble requires a tremendous amount of "luck." Most Boltzmann bubbles compatible with our current evidence $E$ are not produced by a large fluctuation. The overwhelming majority of them are in

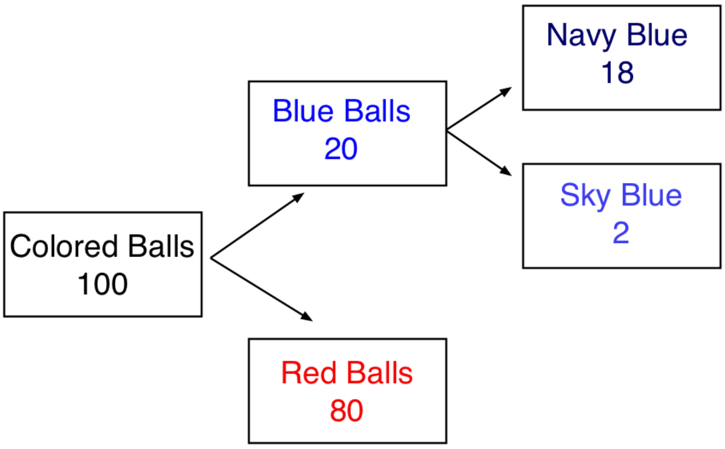


Figure 6: Analogy to an urn problem

fact produced by the smallest fluctuations that dip from equilibrium down into $E$ and then back up to equilibrium. Hence, if we compare MEH′ with the following hypothesis, it is difficult to (epistemically) justify the assignment of significant credence to it:

**MEH$^{\text{14 billion years}}$:** We are currently located in a *normal macrostate* produced by a large fluctuation whose entropy minimum is about 14 billion years away from us; we are equally likely to be any observers (inside such macrostates) that have our current experiences.

If MEH′ and MEH$^{\text{14 billion years}}$ are both empirically adequate, then the relevant question is how we should assign our priors. Given that there are way more medium-level fluctuations than large fluctuations that dip into a low-entropy macrostate, it seems that we should place much more credence in MEH′ than in MEH$^{\text{14 billion years}}$. To illustrate this intuition with an analogy, consider an urn problem drawn in Figure 6. Suppose there are 100 colored balls in the urn and we are going to draw one at random. Suppose for some reason we know that the ball we draw is not red, then it will be a blue ball. Hypothesis 1 says its color is navy blue; Hypothesis 2 says it is sky blue. It seems that we should have way more credence in the first hypothesis than the second hypothesis. In §3.5, we return to the question of empirical adequacy of MEH′ and see more clearly the relevance of the analogy with the urn problem. But for now, we assume that it is compatible (and coheres with) our evidence. Let us thus return to $T_F^*$, the combined theory of $T_F + MEH'$.

## 3.3 The Master Argument*

We are ready to show that $T_F^*$ and $T_M^*$ are on par with respect to evidence $E$. We can appeal to Bayesian reasoning and make the following argument, a revised version of the Master Argument:

**The Master Argument***

P1 Our current evidence $E^*$ is as likely on $T_F^*$ as on $T_M^*$:

$$P(E^*|T_F^*) = P(E^*|T_M^*).$$

P2 $T_F^*$ is roughly as intrinsically likely as $T_M^*$:

$$P(T_F^*) \approx P(T_M^*).$$

C $T_F^*$ is roughly as likely as $T_M^*$ given our current evidence $E^*$:

$$P(T_F^*|E^*) \approx P(T_M^*|E^*).$$

The Master Argument* is valid by an application of Bayes's theorem (assuming $P(E^*) \neq 0$):

$$\frac{P(T_F^*|E^*)}{P(T_M^*|E^*)} = \frac{P(T_F^*)}{P(T_M^*)} \times \frac{P(E^*|T_F^*)}{P(E^*|T_M^*)}.$$

First, P1 follows from the statements of NPH in $T_M^*$ and MEH' in $T_F^*$. The only macrostates allowed (to be the current one) by $T_F^*$ coincides with those allowed by $T_M^*$. The ratio between macrostates compatible with $E$ and all possible macrostates is exactly the same on the two theories. Hence, the probabilities they assign to our current evidence $E^*$ are the same.

Second, there are good reasons to believe that P2 is true. I assume here that a simpler and less *ad hoc* theory will be more intrinsically likely than one that is more complex and *ad hoc*. These are delicate matters of judgment. We will only argue that the two theories are of the same order of simplicity and *ad hocery*, and that they are roughly as intrinsically likely as each other. On the face of it, MEH' in $T_F^*$ seems highly *ad hoc*. Given the possibility of so many medium-entropy fluctuations, why choose only the ones compatible with NPH? It seems that we have engineered the result by putting it into the theory by hand. However, a similar question can be asked of NPH in $T_M^*$. Given the possibility of locating ourselves in so many different epochs, why choose only the first epoch to be where we can be? So it seems that NPH is equally suspect. Thus, NPH and MEH' may be equally *ad hoc*. Hence, they seem to be tied in this respect.

There is another respect that may be relevant to intrinsic probability. NPH is formulated without reference to MEH', but MEH' is formulated with explicit reference to NPH. The definition of *normal macrostates* invokes NPH. Thus, MEH' may seem more *extrinsic* than NPH, in the sense that it exploits the success of another theory. But it is not clear to me if extrinsicness is a bad thing in this case. If NPH provides a simple way to state the restriction to certain medium-entropy macrostates, then it seems that we can and should use that fact in formulating $T_F^*$.[25] In any case, we only need to show that the two theories are roughly equal in intrinsic probability. Even if extrinsicness knocks out some points for $T_F^*$, as long as the disadvantage is not decisive, the two theories are nonetheless of the same level of intrinsic probability.

Hence, we have good reasons to accept the conclusion that $T_F^*$ is roughly as likely as $T_M^*$ given our current evidence $E$.

## 3.4 Skeptical Consequences

If we accept the conclusion of the Master Argument, then we are in trouble. Suppose that $T_F^*$ and $T_M^*$ are the only two theories currently under consideration. Given the conclusion of the Master Argument, our credence in each theory should be roughly 0.5.[26]

As we have explained earlier, while $T_M^*$ predicts a normal past history, $T_F^*$ predicts a radically different one. According to $T_F^*$, most likely we are in a medium-entropy fluctuation. Our future is normal: entropy will increase and things will appear older. However, our past is very different from what we remembered. In fact, we looked not younger, but older, five years ago, since the entropy was higher the further we go into the past. The past will not be like a normal past, given that we connect a normal past with the appearance of younger selves.

$T_F^*$ predicts a symmetric history, one with entropy growing to the past and to the future. This means that our current records and memories about the past are systematically false. My photograph of a five-year old person, with fewer wrinkles and more hair, does not resemble anyone in my past. The person in my past in fact has more wrinkles and fewer hair. In this way, most of our records about the past, which indicate a lower-entropy past, are false. And most of our memories and beliefs about the past are false. The records and memories about the past did not grow from a Big Bang state. Rather, they come from a random fluctuation out of thermal equilibrium, creating the *impression* of a low-entropy past.

If our credence in $T_F^*$ should be roughly 0.5, then our credence in the following should be roughly 0.5:

**Skepticism about the Past:** Most of our beliefs about the past are false.

[25] We have other examples of such extrinsic exploitation in statistical mechanics. For example, an individualistic Boltzmannian theory of statistical mechanics may exploit the success of statistical ensembles developed by the Gibbsian theory. The statistical ensembles are not fundamental in the Boltzmannian theory, but they are nonetheless useful in simplifying the descriptions and calculations of various thermodynamic quantities.

[26] To be more realistic, we should consider more options, for example the possibility that both theories are false. That may well be the case. However, given that these theories have had high explanatory success and that we do not have any concrete proposal that does better, we should still assign some significant credences to $T_F^*$ and $T_M^*$. To be sure, someone who is convinced of the pessimistic meta-induction argument will not be troubled by the skeptical consequences (or perhaps any conclusion of significance that we draw from contemporary science).

But this is not the full scope of the trouble we are in. Many of our beliefs about the present and about the future are partially based on our beliefs about the past (via induction and memory). For example, I believe that tigers are dangerous, based on what I learnt about tigers and biology many years ago. But the learning experience probably did not happen. I believe that the UK will most likely exit the European Union, based on what I have read in newspapers and online articles. But the reading probably did not happen. My memories may be created by a random fluctuation.

If we come to have around 0.5 credence that most of our beliefs about the past are false, then we should accordingly adjust our credences about the present and the future. Hence, we should significantly lower our credences in many of our beliefs about the present and the future. We thus enter into a state of agnosticism and doubt about many beliefs that are dear to us. That is a kind of radical epistemological skepticism. The skeptical problem persists even if $T_F^*$ and $T_M^*$ are not the only hypotheses under consideration. If we also consider $T_F$ + MEH$^{\text{14 billion years}}$, the overall probabilistic boost to counter skepticism will be extremely small, since our prior distribution would assign a very low probability on MEH$^{\text{14 billion years}}$ as a large fluctuation that produces a 14-billion year Boltzmann bubble is extremely rare.

How should we navigate the world if we are convinced of such an argument? I do not yet have an answer. It seems to be a surprising, if not paralyzing, lesson to draw from physics. Physics has challenged many of our preconceived beliefs, such as about solidity, space and time, the microscopic reality, and possibility of actual universes outside our own. However, the skeptical consequences we draw from statistical mechanics, and in particular from theories that attempt to explain the thermodynamic arrow of time, just seem too much. Should we find new physical theories that make sure that the skeptical conclusions do not follow? Or should we embrace the surprising consequences as just another conceptual revision required by physics? Different people will have different judgments here.

Carroll (2017) seems to suggest the first strategy. In the context of discussions about Boltzmann brains in cosmological theories, he recognizes that the prevalence of Boltzmann brains threatens the epistemic status of the theory. He suggests that we try to find cosmological models in which typical observers are normal people and not Boltzmann brains. The same can be said about Boltzmann bubbles: one could require that cosmological models validate the principle that typical observers do not live in a Boltzmann bubble but have normal history. But again, it is not clear how to define "normal history" and it is perhaps vague where the boundary is.

Perhaps there are other strategies beyond the above two. In the next section, I will discuss three different strategies.

## 3.5 Empirical Incoherence?

An initially promising way to respond to the skeptical argument is to point out that $T_F^*$ is *empirically incoherent*: the theory undermines the empirical evidence we have for accepting it in the first place. Presumably, we came to consider $T_F^*$ based on empirical evidence for statistical mechanics and cosmology. However, $T_F^*$ now

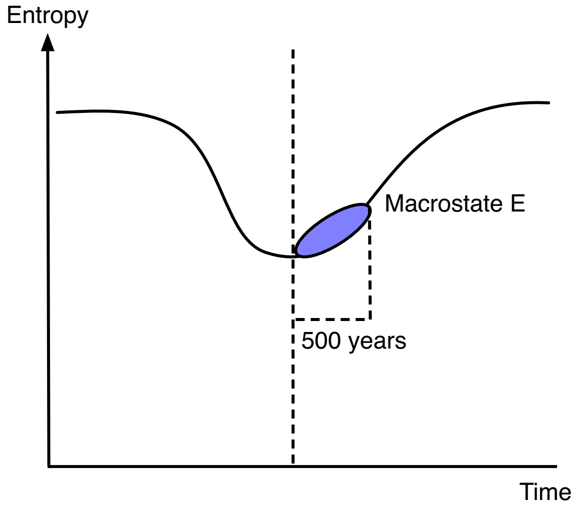


Figure 7: A Boltzmann Bubble that extends 500 years.

predicts that our evidence for accepting it has a significant probability ($\approx 0.5$) to be false. Our memories and records for past experiments and observations might well be the results of random fluctuations from equilibrium. Hence, we have good reasons to reject $T_F^*$.[27]

However, it is not clear if $T_F^*$ has to be empirically incoherent. Suppose the theory is supported by evidence $E$ collected in 500 years of normal history. This period includes all the experiments, observations, and derivations we made for classical mechanics, quantum mechanics, cosmology, and statistical mechanics. That is, we allow $E$ to include not only the current macrostate but also macrostates that stretch to 500 years in the past. Such $E$ can still be produced by medium fluctuations, but we are not longer near the minimum but 500 years away from it. We can revise MEH' as follows:

**MEH$^{\textbf{500 years}}$:** We are currently located 500 years away from the minimum of a medium fluctuation that produces a *normal macrostate*; we are equally likely to be any observers (inside such macrostates) that have our current experiences.

$T_F^*$ would predict that our memories about all the great physics experiments and observations since the scientific revolution were veridical: they really happened. What happened in the middle ages were completely different from what we thought we know, but that does not interfere with the reasons we have for accepting the scientific theory. The Boltzmann bubble which contains us stretches both in space and in time. (See Figure 7.)

In other words, we can let our evidence $E$ to include whatever is necessary to support accepting $T_F^*$. If that requires 500 years of "normal" history, then make the

[27]See Barrett (1996, 1999) for discussions about empirical incoherence, especially in the context of quantum theories.

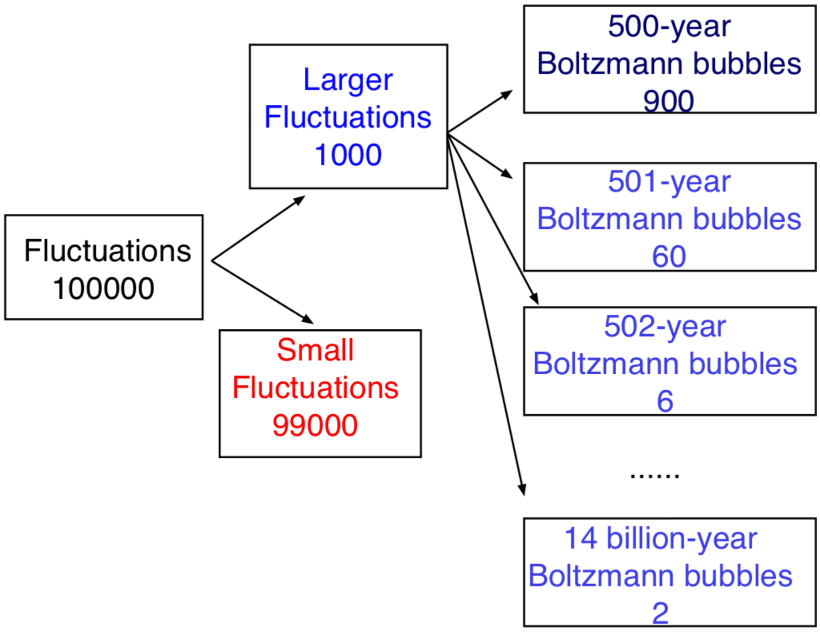


Figure 8: A rough comparison of the numbers of different fluctuations (the numbers are illustrative only and do not reflect the relevant scale)

macrostate $E$ stretch back to 500 years. That is still compatible with accepting $T_F^*$, which predicts that we most likely live in a Boltzmann bubble and we are 500 years from the minimum of the recent fluctuation.

In fact, MEH$^{500 \text{ years}}$ makes $E$ more likely than NPH does, since for each medium fluctuation there are two possibilities for us: either on the "left" of the minimum or the "right." This extra probabilistic boost could be significant. For example, it could make even (in posterior probability) any loss to intrinsic probability in $T_F^*$ after we change MEH′ to MEH$^{500 \text{ years}}$.

But if we are happy to accept MEH$^{500 \text{ years}}$ as a modification of $T_F^*$, why not go further? Why not accept MEH$^{501 \text{ years}}$, MEH$^{502 \text{ years}}$, ..., or even MEH$^{14 \text{ billion years}}$? After all, the longer we stretch the Boltzmann bubble, the more normal history there will be in our specific fluctuation. Other things being equal, wouldn't it be nice to have more normal history? The answer was hinted already in §3.2. It would indeed be very nice, but I worry whether that is epistemically the right thing to do. After all, there are overwhelmingly more fluctuations compatible with the selections in MEH$^{500 \text{ years}}$ than in MEH$^{501 \text{ years}}$. And in fact, it is probably the case that there are overwhelmingly more fluctuations compatible with the selections in MEH$^{500 \text{ years}}$ than all the fluctuations compatible with either MEH$^{501 \text{ years}}$, MEH$^{502 \text{ years}}$, ... , or MEH$^{14 \text{ billion years}}$. It seems that we should not have higher credence in the disjunction of MEH$^{501 \text{ years}}$, MEH$^{502 \text{ years}}$, ... and MEH$^{14 \text{ billion years}}$ than in MEH$^{500 \text{ years}}$. In Figure 8, we provide a rough comparison of the numbers of different fluctuations. If for *a priori* reasons (such as to avoid incoherence) we need to postulate our temporal location in larger fluctuations (much larger than Boltzmann brain fluctuations, and have at least 500 years of normal history), we are still left with some uncertainty over the size of fluctuation we are in. There are much more 500-year Boltzmann bubbles than 501-year Boltzmann bubbles, 502-year Boltzmann bubbles, and 14-billion-year Boltzmann bubbles. Of course, the comparative difference gets exponentially larger as we stretch the Boltzmann bubble from 500 years to 14 billion years (since fluctuations have to be extremely delicate to dip all the way to something like a Big Bang state).

By the same reasoning, should we have most of our credence in just MEH′ since it is compatible with much more fluctuations than $\mathrm{MEH}^{500\text{ years}}$? No, the empirical incoherence of MEH″ is sufficient to subtract much of our credence in that hypothesis. In contrast, we assume that given 500 years of normal history, the fluctuation hypothesis would be empirically coherent.

### 3.6 Other Responses

*(A) Disputing the Priors.* Another way to respond to the Master Argument is to point out that we should choose priors that overwhelmingly favor $T^*_M$. This is because if we give some weight to $T^*_F$, then the previous argument will lead us to skepticism. To lead a successful epistemic life, we ought to choose priors that do not lead us into skepticism. For example, we can assign very low prior probability to $T^*_F$ to block P2 of the Master Argument*.

This is what I would like to do in practice. However, is it always epistemically justified? Can we always sweep under the rug any skeptical conclusion we do not like? It would be helpful to have more general principles to guide us here.

In so far as there are any objective norms for credences, I would think that they favor simpler theories. The reason we are warranted to assign extremely low priors to skeptical hypotheses, in many cases, is because the skeptical hypotheses are highly complex. The Evil Demon Hypothesis, the Brain-in-Vat Hypothesis, and the Dream Hypothesis, one could argue, are much more complex than the Real World Hypothesis, if we spell them out in detail. The Real World Hypothesis can be described with ordinary simple laws of physics plus the usual initial conditions while the skeptical hypotheses have to be supplemented with extra details that produce the skeptical scenarios. (The Boltzmann Brain Hypothesis may be simpler than $T^*_F$, but the Boltzmann Brain Hypothesis is not empirically adequate with respect to our current evidence $E$, as we take $E$ to be quite generous and externalistic.)

In contrast, $T^*_F$ is on the same level of simplicity as $T^*_M$: they are similar cosmological theories with *de se* postulates. If we are warranted to assign low credences in $T^*_F$, then it must come from other considerations beyond the usual ones based on the complexity of the skeptical theories. However, if such considerations do apply, after careful examination of the case at hand, that would be a welcome result indeed.

*(B) Fluctuations are impossible in our universe.* Although we have assumed at the outset of the paper that fluctuations are physically possible, it is nonetheless interesting to consider the strategy that denies that. If there are no fluctuations, then Boltzmann bubbles cannot form. The Fluctuation Hypothesis would not work. And maybe there are not even Boltzmann brains. In such a universe, Nature is kind to us. The kind of empirical underdetermination would not hold and skeptical

consequences would not arise (at least not by our arguments).

However, how much confidence should we assign to the physical impossibility of fluctuations? It is unclear as it is still not settled in physics. But suppose we have reasonable confidence (say, 0.6) in such an impossibility. Then there is still some probability (say, 0.4) that fluctuations can lead to the kind of underdetermination and skeptical worries. In that case, the influence of skeptical hypothesis is much lower (knocking out at most 0.2 credence in many beliefs). However, the influence will still be felt, and a weaker kind of skeptical worry will arise (for many propositions that have threshold credences around 0.5). Thus, it will still be interesting to resolve the skeptical worry on the assumption that fluctuations are possible.

In summary, the three responses (empirical incoherence, disputing the priors, and denying physical possibility) to the skeptical argument are initially promising but in the end inconclusive.

## 4 Conclusion

A long-standing problem in the foundations of physics has been to explain the origin of the arrow of time. One promising explanation is offered by the Mentaculus Theory. However, it faces a *de se* problem if there are many copies of us in the distant past or the distant future. In the first part of this paper, we argued that we need to add *de se* probabilities to save the Mentaculus Theory and arrive at a monotonic increase of entropy in our epoch of the universe. In the second part of the paper, we show that such a move opens a new door to skepticism. I show this by using *de se* probabilities to revive the Fluctuation Theory. I argue that our evidence underdetermines between the two theories after we include certain postulates about *de se* probabilities. We are thus led to epistemological skepticism.

The worries we raise here can be construed as new objections to the Mentaculus Theory and to the Past Hypothesis explanation. However, I hope to offer these as friendly invitations for defenders of $T^*_M$ to clarify and develop the philosophical principles underlying their explanation. What is particularly interesting is that a skeptical problem naturally arises when we put together time's arrow with *de se* probabilities. It would be worth thinking about how various responses to skepticism bear on this particular instance. I am optimistic that such responses can be found. After all, the Boltzmannian project in general and the Mentaculus theory in particular have led to undeniably impressive progress in our understanding of the arrows of time. At the moment, however, it is not an exaggeration to say that we do not completely understand the logic behind our belief that we do not live in a fluctuation. Even if we are not led to epistemic skepticism by the previous arguments, it seems that we should at least embrace epistemic modesty about *de se* probabilities, time's arrows, and the scientific project of explaining our knowledge of the past, despite the impressive progress we have made.

## Acknowledgement

I am grateful for helpful discussions with David Albert, Ben Bronner, Barry Loewer, David Wallace, Tim Maudlin, Sean Carroll, David Chalmers, Antony Eagle, Bob Beddor, Isaac Wilhelm, Veronica Gomez, Sheldon Goldstein, Cameron Kirk-Giannini , Kelsey Laity-D'Agostino, Joel Lebowitz, Branden Fitelson, Daniel Rubio, David Black, Zee Perry, Richard Pettigrew, Karim Thébault, Peter Graham, Eric Schwitzgebel, Charles Sebens, Christopher Hauser, John Hawthorne, Michael Huemer, Bradley Monton, Craig Callender, Christopher Hitchcock, Oli Odoffin, Jonathan Schaffer, Miriam Schoenfield, Katie Steele, Michael Tooley, Alan Hájek, as well as audiences at the 3rd Boulder Formal Value Workshop, CalTech HPS Lecture Series, Bristol Center for Philosophy of Science, the 2017 AAP Conference at Adelaide, the College of William and Mary, Wuhan University, UCSD philosophy of science group, Rutgers Philosophy of Science Group, and Rutgers Cognitive Science Grad Talk Series.

## References

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

Albert, D. Z. (2015). *After physics*. Harvard University Press.

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

Barrett, J. A. (1999). *The quantum mechanics of minds and worlds*. OUP Oxford.

Boltzmann, L. (2012). *Lectures on gas theory*. Courier Corporation.

Callender, C. (2004). Measures, explanations and the past: Should 'special' initial conditions be explained? *The British journal for the philosophy of science*, 55(2):195–217.

Callender, C. (2011). Thermodynamic asymmetry in time. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2011 edition.

Carroll, S. (2010). *From eternity to here: the quest for the ultimate theory of time*. Penguin.

Carroll, S. M. (2017). Why boltzmann brains are bad. *arXiv preprint arXiv:1702.00850*.

Carroll, S. M. and Chen, J. (2004). Spontaneous inflation and the origin of the arrow of time. *arXiv preprint hep-th/0410270*.

Clausius, R. (1867). *The mechanical theory of heat: with its applications to the steam-engine and to the physical properties of bodies*. J. van Voorst.

Coen, E. and Coen, J. (2010). *A serious man*. Faber & Faber.

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.

Feynman, R. P., Leighton, R. B., and Sands, M. (2011). *The Feynman lectures on physics, Vol. I: The new millennium edition: mainly mechanics, radiation, and heat*, volume 1. Basic books (first published in 1964 by Addison-Wesley).

Fitelson, B. (2006). Does the principle of indifference rest on a mistake? *Unpublished manuscript.*

Fitelson, B. (2007). Likelihoodism, bayesianism, and relational confirmation. *Synthese*, 156(3):473–489.

Goldstein, S. (2001). Boltzmann's approach to statistical mechanics. In *Chance in physics*, pages 39–54. Springer.

Goldstein, S. and Tumulka, R. (2011). Approach to thermal equilibrium of macroscopic quantum systems. In *Non-Equilibrium Statistical Physics Today: Proceedings of the 11th Granada Seminar on Computational and Statistical Physics, AIP Conference Proceedings*, volume 1332, pages 155–163. American Institute of Physics, New York.

Goldstein, S., Tumulka, R., and Zanghì, N. (2016). Is the hypothesis about a low entropy initial state of the universe necessary for explaining the arrow of time? *Physical Review D*, 94(2):023520.

Ismael, J. (2008). Raid! dissolving the big, bad bug. *Noûs*, 42(2):292–307.

Lazarovici, D. and Reichert, P. (2018). Arrow of time without a past hypothesis. *arXiv preprint arXiv:1809.04646*.

Lebowitz, J. L. (2008). Time's arrow and boltzmann's entropy. *Scholarpedia*, 3(4):3448.

Lewis, D. (1979). Attitudes *De Dicto* and *De Se*. *Philosophical Review*, 88:513–43.

Loewer, B. (2001). Determinism and chance. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 32(4):609–620.

Loewer, B. (2004). David lewis's humean theory of objective chance. *Philosophy of Science*, 71(5):1115–1125.

Loewer, B. (2016). The Mentaculus Vision. In Barry Loewer, Brad Weslake, and Eric Winsberg (eds.) *Time's Arrow and World's Probability Structure*. Harvard University Press, forthcoming.

Manson, N. A. and Thrush, M. J. (2003). Fine-tuning, multiple universes, and the "this universe" objection. *Pacific Philosophical Quarterly*, 84(1):67–83.

McGrew, T., McGrew, L., and Vestrup, E. (2001). Probabilities and the fine-tuning argument: a sceptical view. *Mind*, 110(440):1027–1038.

North, J. (2010). An empirical approach to symmetry and probability. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 41(1):27–40.

North, J. (2011). Time in thermodynamics. *The oxford handbook of philosophy of time*, pages 312–350.

Penrose, R. (1979). Singularities and time-asymmetry. In Hawking, S. and Israel, W., editors, *General relativity*, pages 581–638.

Penrose, R. (1989). *The emperor's new mind: concerning computers, minds, and the laws of physics*. Oxford University Press.

Perry, J. (1979). The problem of the essential indexical. *Noûs*, pages 3–21.

Pettigrew, R. (2014). Accuracy, risk, and the principle of indifference. *Philosophy and Phenomenological Research*.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Pettigrew, R. and Titelbaum, M. G. (2014). Deference done right. *Philosophers' Imprint*.

Price, H. (1997). Cosmology, time's arrow, and that old double standard. *Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time*, pages 66–96.

Sebens, C. T. and Carroll, S. M. (2016). Self-locating uncertainty and the origin of probability in everettian quantum mechanics. *The British Journal for the Philosophy of Science*, 69(1):25–74.

Uffink, J. and Valente, G. (2010). Time's arrow and lanford?s theorem. *Le temps*, pages 141–173.

Wallace, D. (2011). The logic of the past hypothesis.

White, R. (2000). Fine-tuning and multiple universes. *Nous*, 34(2):260–276.

Winsberg, E. (2012). Bumps on the road to here (from eternity). *Entropy*, 14(3):390–406.