**Atheistic Induction by Boltzmann Brains**

**Bradley Monton, Wuhan University**

**forthcoming in *Two Dozen (or so) Arguments for God: The Plantinga Project,* eds Jerry L. Walls and Trent Dougherty (2018)**

**Abstract**

I present a new thermodynamic argument for the existence of God. Naturalistic physics provides evidence for the failure of induction, because it provides evidence that the past is not at all what you think it is, and your existence is just a momentary fluctuation. The fact that you are not a momentary fluctuation thus provides evidence for the existence of God – God would ensure that the past is roughly what we think it is, and you have been in existence for roughly the amount of time you think you have. I don’t have a definitive way for the atheist to refute this argument, but I give one suggestion that relies on physics-based simplicity considerations. I close with an epistemological discussion of self-undermining arguments.

**keywords**

thermodynamics, argument for the existence of God, induction, simplicity, naturalism, atheism, past hypothesis, Boltzmann brains, problem of induction, modal realism, self-undermining arguments

**1. Introduction**

Does the fact that our inductive practices are successful provide evidence for the existence of God? Do simplicity considerations provide such evidence? I will mostly focus on the argument from induction, and in doing so I will present a new thermodynamic argument for the existence of God. But let’s begin with an opening salvo on simplicity.

With regard to the argument from simplicity for the existence of God, here’s a natural question to ask: is the hypothesis that God exists a very simple hypothesis (as for example Richard Swinburne (1979) claims), or is the hypothesis that God exists a very complex hypothesis (as for example Richard Dawkins (2008) claims)? If simplicity is a guide to truth (as both those Richards think it is), then by Swinburne’s lights, the prior probability for the God hypothesis is high, while by Dawkins’ lights, the prior probability for the God hypothesis is low. And according to them, after one takes into account all the evidence to get a posterior probability, their posteriors aren’t much different than their priors – the simplicity judgments are doing the bulk of the work.

The question of whether simplicity is a guide to truth is a vexing one. It is natural to think that simplicity is a guide to truth in the context of, for example, curve-fitting in physics. Suppose one has a finite set of data points representing the relationship between two parameters in physics, and one wants to draw a curve that projects a relationship from those finite data points to the values of the parameters in-between. The problem is that there are an infinite number of curves that can be drawn which intersect all the data points. Most of these curves are unsimple (in a crazy, all-over-the-place sort of way), and hence are rejected by physicists. And, as more data points are gathered, typically the simpler curves turn out to make the right predictions for those new data points. It sure looks like simplicity is a guide to truth.

But philosophers are sometimes too willing to take these relatively uncontroversial judgements from physics and wildly extrapolate them into controversial judgements in philosophy. For example, consider ontological parsimony (parsimony of objects) versus ideological parsimony (parsimony of concepts). Is it ontological or ideological parsimony that is a hallmark of simplicity and hence of truth? While some philosophers attempt to provide an objective answer to this question (see for example Cowling 2013), I am skeptical. Maybe God exists and God has an opinion, but absent God’s judgement, I maintain that it’s just a matter of personal preference which counts as simple, and that neither is a hallmark of truth.

My plan in this essay is to focus on induction, and to only bring in simplicity considerations when they are well-grounded in physics. You might then think that such physics-based simplicity considerations wouldn’t come into play in an essay on God and induction. But you might also recognize that the previous sentence constitutes foreshadowing that they will.

So, back to induction – here is the schedule of this essay. One can raise worries about the viability of induction generally, as Hume did – my discussion of that will be Preliminary Match #1. One can also raise worries about the viability of induction as applied to a particular philosophical theory – I’ll discuss that in the context of David Lewis’s modal realism in Preliminary Match #2. Finally, we’ll turn to the Main Event, on how naturalistic physics provides positive evidence for the failure of induction, and how to potentially overcome that, whether naturalistically or theistically. (And stick around for the epistemological epilogue, on self-undermining arguments.)

**2. Preliminary Match #1: The General Problem of Induction**

Hume famously argued that we have no grounds for believing that our inductive practices will be reliable in the future. At least, we have no grounds for thinking this, assuming naturalism. As for example Descartes showed, one can get around skeptical worries by establishing (or presupposing) the existence of God, and then pointing out that God is not a deceiver. God wouldn’t deceive us, hence our inductive practices will be reliable in the future.

Well, that’s nice that believing in God provides a solution to Hume’s problem of induction. But wishes aren’t horses – can one turn this into an argument for theism? Let’s consider the following transcendental argument:

*Premise 1*: A precondition for reliable inductive reasoning is the truth of theism.

*Premise 2*: We do reliably engage in inductive reasoning.

*Conclusion*: Theism is true.

(Bare theism wouldn’t do: it would take something more, like non-deceiver theism. I’ll henceforth assume that as built into the theistic hypothesis.)

The above argument has the drawback of being unsound. The crucial transcendental premise is false; it’s certainly *possible* that we could engage in reliable inductive reasoning in a naturalistic world. All Hume pointed out is that we have no reason to *expect* our inductive reasoning to be reliable.

Perhaps the argument should be put in terms of our justification for inductive reasoning:

*Premise 1*: A precondition for us being justified in engaging in inductive reasoning is the truth of theism.

*Premise 2*: We are justified in engaging in inductive reasoning.

*Conclusion*: Theism is true.

But the (or at least a) problem with this new argument is that Premise 2 is questionable, and question-begging. We don’t know that we are justified in engaging in inductive reasoning – we’ve been successful at it in the past, but that doesn’t follow that we are epistemologically justified in that practice.

Instead of giving a transcendental argument, perhaps one can instead give an inductive argument:

*Premise 1*: Under the assumption of naturalism, we have no reason to expect our inductive practices to be reliable.

*Premise 2*: Under the assumption of theism, we do have reason to expect our inductive practices to be reliable.

*Premise 3*: Our inductive practices are reliable.

*Conclusion*: The reliability of our inductive practices provides evidence for the truth of theism over naturalism.

That argument, admittedly, has some inductive strength – but, I maintain, not much strength. To see why, compare that argument to this one:

*Premise 1*: Under the assumption of naturalism, we have no reason to expect our next coin flip to land heads (rather than tails).

*Premise 2*: Under the assumption that a supernatural sprite who loves heads has just popped into existence in the room, we do have reason to expect our next coin flip to land heads.

*Premise 3*: Our next coin flip lands heads.

*Conclusion*: That coin flip landing heads provides evidence for the existence of the supernatural sprite and against naturalism.

I am willing to concede that, from a probability theory standpoint, that sprite argument works. But my prior probabilities are such that the level of support the argument provides for the sprite hypothesis is minimal. And the same holds for the previous argument for theism. Sometimes, things are the case, even if we have no reason to expect them to be the case, and that shouldn’t provide much support for the existence of a supernatural entity. This holds for the coin landing heads, and this also holds for the hypothesis that our inductive practices are reliable.

Here’s another way to look at this. There are some things theists can explain that naturalists can’t. But naturalists should be comfortable with this. Most naturalists would be happy to concede that they don’t have an explanation of why there is something rather than nothing, why the universe has the laws of nature that it does, and so on. Theists have an explanation for all this: God created something, God created the laws, and so on. Similarly, naturalists can’t explain why our inductive practices are reliable; theists can. It doesn’t follow that the theist’s explanation is true.

**3. Preliminary Match #2: The Modal Realist’s Problem of Induction**

Above, I discussed a situation where we have no reason to think that induction is reliable, under the assumption that naturalism is true. But what if we found ourselves in a situation where we have positive reason to think that induction is unreliable, under the assumption that naturalism is true? I’ll now show that there is at least one important philosophical theory that has this consequence. One way to avoid this consequence, of course, is to reject naturalism. My preferred method is to reject the important philosophical theory in question.

 The theory I have in mind is David Lewis’s modal realism. Lewis (1986) believes that there are an infinite number of physical worlds, spatiotemporally disconnected from each other, representing every possible way that the world could be. Lewis also believes in a principle of recombination, according to which “patching together parts of different possible worlds yields another possible world” (1986, 87–88). Consider all the worlds with a history like ours, up to the present moment. By Lewis’s lights, *any* possible future can be patched together with our past, and there is a world in existence with that past and that future.

So think about all the possible worlds with a history like ours, and the wide variety of futures they have. There are, in an intuitive sense at least, a comparatively small number of worlds where things go as expected by the lights of inductive reasoning, and a comparatively large number of worlds that they haywire by the lights of inductive reasoning. So according to Lewis’s theory, we should reason counter-inductively: we should think it very unlikely that the principle of induction will continue to hold in the future. Just think about all the possible parts of worlds there are: a part of a world with just one electron, a part of a world with just two electrons, a part of a world with just one dragon, et cetera. Any of these parts could be the complete future of our world, by Lewis’s lights. So, it’s vastly more likely that the future will not line up nicely with the past than that it will.

Interestingly, Plantinga suggests otherwise. He writes:

there are plenty of possible worlds (worlds run by mischievous Cartesian demons, perhaps) in which things go just as they have up to the present, but then go completely crazy; there are as many worlds like that as worlds in which induction will continue to be a reliable source of belief. (Plantinga 1993, 125)

While Plantinga isn’t specifically discussing Lewis in this context, he is discussing a possible-worlds framework that (for our purposes at least) is compatible with Lewis’s framework. So in this context, I respectfully maintain that Plantinga is mistaken. I maintain there aren’t “as many” worlds where induction is successful as not; instead I maintain that there are *many fewer* worlds where induction is successful as not. We have positive reason to think that, by Lewis’s lights, the future will most likely go completely crazy.

To really show that Plantinga is mistaken, though, one would have to have a measure over the space of possible worlds, which specifies for any set of worlds how likely they are compared to another set. While we often make intuitive judgements that yield such results, we have no specification of what an overall measure would be. (An example of an intuitive judgement: worlds with a history like ours where every human gets struck by lightning in the next hour are less likely than worlds with a history like ours where some but not all humans get struck by lightning in the next hour.)

Let’s assume that I’m correct that there are a comparatively small number of worlds where things go as expected by the lights of inductive reasoning, and a comparatively large number of worlds where they go as unexpected, and let’s assume that the measure over the worlds treats all these worlds equiprobably. Then if one believed in modal realism (as some of us do) and found oneself in a world with a particular history (as all of us do) and contemplated whether inductive reasoning is going to hold in the future (as you are so contemplating, now that I’ve said it), one would recognize that the vast majority of worlds with this history are going to go completely differently in the future, while only rare worlds with this history will have a future that conforms with our inductive reasoning, and one should be incredibly scared. At least, one should be as scared as one would be seeing the rampaging pirate with a sword coming at one’s neck – in either scenario one is probably not long for this world. The vast majority of possible worlds with a history like ours are ones where the immediate future does not contain you or me or any of the things we care about. Let us contemplate this while we can, for in the next moment we most likely die.

Perhaps you are not happy with this result. How should you reject it? One solution – my favored solution – is to reject modal realism altogether. Another solution is to postulate a theistic version of modal realism: only those possible worlds exist where God exists, and God is not a deceiver, so induction is reliable in all possible worlds. Either move, of course, constitutes a rejection of David Lewis’s version of modal realism.

Let’s step back. I have been raising the worry of inductive skepticism about the future. But I could have presented a parallel skeptical problem about the past. By the same principle of recombination, one can combine any past history with a present like ours. So just as we should doubt our inductive reasoning about the future given Lewis’s modal realism, so we should doubt our memories and inferences about the past.

But let’s not harp on modal realism more. Let’s move on to the Main Event, in which I’ll show that our best physics suggests that in the actual world we are facing just such a skeptical problem, about both the past and the future.

**4. The Main Event: The Naturalistic Implausibility of the Past Hypothesis**

***4.1 Boltzmann Brains***

Consider a regular raw chicken egg, balanced precariously on the edge of a table. The egg eventually falls onto the floor, smashing into many goopy pieces. There is nothing unfamiliar about such an event. But the laws of physics are (as far as physicists can tell) time-reversal invariant (with minor exceptions which, as far as physicists can tell, wouldn’t have any relevance to everyday events like egg-fallings; I’ll ignore these minor exceptions henceforth). This means that, in principle, the time-reverse of the story I just told is also physically possible. Minor undulations in the floor could congregate on a particular spot, pushing the pieces of the egg up into the air in such a way that the pieces form into a whole egg and the egg lands on the edge of the table. This is – uncontroversially – physically possible. So why do we never see it?

 The reason we don’t is that it’s incredibly unlikely to occur – it’s incredibly unlikely that forces would align in just such a way as to reform a broken egg. This incredible improbability is captured in the second law of thermodynamics – that entropy in a closed system never decreases. The broken egg on the floor is a higher-entropy state than the whole egg on the table, so by the second law the broken egg could never reform into the whole egg on the table. But if laws are understood as exceptionless regularities, then the second law isn’t really a law; it’s just a statistical generalization. The generalization holds because, when considering all the possible states of a system, the low-entropy states are incredibly rare, and high-entropy states are incredibly likely. For a given macro-description of a system (like “broken egg on the floor”) the number of ways the system evolves where entropy stays the same or increases vastly outnumbers the number of ways the system evolves where entropy decreases. (To be more precise, we’d want a probability measure over the space, but physicists for the most part agree on what the probability measure is, and it yields the result that the low-entropy states are incredibly unlikely.)

 Now let’s think about the universe as a whole. I’m going to give an account, based on standard physics, but with an almost unbelievable conclusion. Once I get to the unbelievable conclusion, I’ll explain how physicists standardly try to avoid it.

Given the current state of the universe, we expect the universe to evolve toward higher-entropy states, since they are vastly more likely than lower-entropy states. This would culminate in what’s called “the heat death” of the universe – particles are spread out throughout the universe (or perhaps clumped due to gravity), but with no structure of the sort that we see in, say, living organisms. Anyway, that’s what we expect for the future – the crucial point I want to make is that, given the time-reversal invariance of the laws of physics, *we should expect the same thing for the past*. Higher-entropy states are vastly more likely than lower-entropy states, so, given the current state of the universe, dynamical trajectories for the universe that get to that current state starting from higher-entropy states are vastly more likely than dynamical trajectories that start from lower-entropy states.

 But if the universe were in a high-entropy “heat death” state, how could one ever get to the low-entropy order that we see around us? Well, physics allows for random fluctuations out of high-entropy states – on rare occasions, a part of the universe will fluctuate into a lower-entropy state, and then return to the high-entropy state. Perhaps that’s what happened to the part that we are in, and we are now on the lower-to-higher entropy trajectory that we expect to be on.

But here’s the rub. The larger the size of the part of the universe that fluctuates into a lower-entropy state, the less likely that is to occur. For you to judge a theory to be viable, it minimally needs to be empirically adequate – it needs to account for your observations. Of the various fluctuations out of a high-entropy state that are compatible with your current observations, the most likely ones are where nothing more than a brain-sized region fluctuates out of the high-entropy state, momentarily generating a lower-entropy organized brain that has exactly the experiences you are having now. But moments ago, that brain didn’t exist, and moments later, it will have ceased to exist – the momentary fluctuation will be over.

This hypothesis that you are such a brain – a “Boltzmann brain” – is compatible with everything that you currently observe. And, strictly speaking, what you currently observe is the totality of your empirical evidence. We sometimes cite what we remember as evidence, but the hypothesis that you are a Boltzmann brain calls into question the reliability of these memories. And, most importantly for our context, the hypothesis that you are a Boltzmann brain calls into question the reliability of your inductive practices – so many of the beliefs you currently have about the future will (within the next few moments, as the system evolves back to the high-entropy state and your brain ceases to exist) turn out to be *false*.

***4.2 The Past Hypothesis***

Let’s step back. In Preliminary Event #1, I reflected on how Hume pointed out that we have no reason to believe that our inductive practices will continue to be successful in the future – but that line of reasoning at least has the happy consequence that we have no reason to believe that they won’t. In Preliminary Event #2, I took up modal realism, and argued that, by the lights of that controversial theory, we do have positive reason to believe that our inductive practices won’t be successful in the future. Now, in the Main Event, I have argued that contemporary physics provides positive reason to believe that your inductive practices won’t continue to be successful – instead you’re just a brain that momentarily fluctuated into existence.

 So how should we avoid this unhappy conclusion? The theist could say that God exists, and God presumably wouldn’t design the world in such a deceptive fashion, and God presumably would value the existence of continuing moral agents, not momentary Boltzmann brains. In short, the theist can evade this Boltzmann-brainian problem of induction.

What would the atheist say? The standard move that atheists make is to postulate *the past hypothesis* – that the universe started in a low-entropy state (or, if the universe didn’t have a temporal beginning, that the past boundary condition of the universe is a low-entropy state).

The past hypothesis is endorsed by many physicists and philosophers of physics (for a nice defense see David Albert (2000, Chapter 4)), but there are dissenters (such as John Earman (2006)). For the purposes of this paper, I will set aside the dissenters, and examine the justifications philosophers have given for endorsing the past hypothesis.

 A problem with the past hypothesis is that it is highly improbable, by the lights of standard physics. It’s way more improbable than the broken egg reformulating and soaring up to the table – it’s as if improbable things like that have happened, but for the universe as a whole.

 Huw Price makes the point vividly:

Suppose … physics had discovered that the matter in the universe is collapsing towards a Big Crunch, fifteen billion years or so in our future – and that as it does so, something very, very extraordinary is happening. … Somehow, by some unimaginably intricate balancing act, the various forces are balancing out, so that by the time of the Big Crunch, matter will have spread itself out with great uniformity. A butterfly – nay, a molecule – out of place, and the whole house of cards would surely collapse!

 As a combination of significance and sheer improbability – the latter judged by well-grounded existing conceptions of how matter might be expected to behave – this discovery would surely trump anything else ever discovered by physics. …

In my view, however, this state of affairs is *exactly* what physics has discovered! I have merely described it in unusual language, reversing the usual temporal conventions … this redescription has no objective significance. If it is a proper matter for explanation described one way, it is a proper matter for explanation described the other. (Price 2002, 115-6)

I agree with Price – it is very, very extraordinary that everything in the universe is lined up in such a way that the universe started with a low-entropy state. I agree that that state is highly improbable, and that that judgement of improbability is well-grounded in physics. And, I agree (pace, for example, Craig Callender (2004)) that such an improbable state cries out for explanation. So what is Price’s explanation? Price (2002, 118) says: “A solution … is not yet at hand. Indeed, it is not yet clear what a solution would look like.”

 One thing bothers me about the approach of Price and all the others who endorse this improbable past hypothesis from within a naturalistic framework. I want to ask them: why reject the Boltzmann brain hypothesis in the first place? The theist has a reason to reject that hypothesis – the theist can point out that God is not a deceiver, and that God values the existence of continuing moral agents. But what reason does the naturalist have, other than wishful thinking that our memories are reliable and our inductive practices will continue to be successful? The naturalist posits the past hypothesis to save those desiderata, but the naturalist should recognize that physics is telling us the hypothesis being postulated to save those desiderata is very, very implausible.

 Many naturalistic endorsers of the past hypothesis don’t even give a reason for rejecting the Boltzmann brain hypothesis – they just take for granted that the hypothesis should be rejected. David Albert (2000, 93-95), to his credit, gives an argument – but unfortunately, as I’ll now show, his argument is flawed.

 Albert first points out, as I did above, that the hypothesis that you are a Boltzmann brain is compatible with everything you currently observe. Indeed, he goes further, and points out that:

there can be nothing at all about the present macrocondition of the world which can possibly count as evidence that the world’s entropy has ever previously been lower. (Albert 2000, 93)

Albert recognizes that you have memories that suggest that entropy was lower – you, for example, looking at the broken egg on the floor, have the memory of it being intact on the table moments before. But, unless we endorse the past hypothesis, physics tells us that it’s overwhelmingly likely that the current state of the world is just a fluctuation from a higher-entropy state, and (hence) it’s overwhelmingly likely that your current memories are nonveridical.

 I agree with all that. But then Albert tries to argue that we do know that the past hypothesis is true:

the fact that the universe came into being in an enormously low-entropy macrocondition cannot possibly be the sort of fact that we know, or ever *will* know, in the way we know of straightforward everyday particular *empirical* facts. We know it *differently*, then. Our grounds for believing it turn out to be more like our grounds for believing general theoretical *laws*. Our grounds (that is) are *inductive*… (Albert 2000, 93-94)

Albert bases his argument on the claim that believing the past hypothesis “turns out to be enormously helpful in making an enormous variety of particular empirical *predictions*.” For example, suppose you dig up a boot from the ground. You would predict that it would be reasonably likely that, if you kept digging, you would find a matching boot. We make predictions like this all the time, and they are often verified. Albert argues that the verification of such predictions is to be expected given the past hypothesis, but there’s no reason to expect them to be verified given that you are a Boltzmann brain. Thus, he believes the past hypothesis.

 The problem with Albert’s argument is that when you think you have made a prediction that has been successfully confirmed, that relies on *memory*. The confirmation cannot come at exactly the same time as the prediction; it’s the nature of prediction that the prediction has to be made first, and then confirmation comes later. So when you are at the time of the confirmation, thinking “oh, my prediction has been confirmed”, you are relying on the memory that you made such a prediction in the first place. But Boltzmann brains can have such memories too – it’s just that such memories for Boltzmann brains are mostly nonveridical. A Boltzmann brain with such memories would have most likely fluctuated into existence with the false belief that it made such predictions, and the false belief that some of these predictions have been confirmed. (Moreover, Boltzmann brains will most likely fluctuate out of existence before they have the chance to confirm any predictions they do really make about the future.) So you might *think* that you’ve made predictions that have been confirmed, but Boltzmann brains would think that too; the fact that you have that thought doesn’t provide evidence that you are not a Boltzmann brain.

 Thus, Albert’s attempt to provide inductive reason that you are not a Boltzmann brain is unsuccessful. Albert has provided no reason for you to think that you really have made successful inductive predictions – and moreover, the evidence from physics is that you are probably a Boltzmann brain and hence you haven’t.

***4.3 Simplicity***

I’ve shown that Albert’s argument for why you should think you’re not a Boltzmann brain is unsuccessful. I have a speculative but potentially better argument to offer the naturalist, based on considerations of simplicity, grounded in physics.

We can get a physics-based definition of simplicity, building on an important new paper by Scott Aaronson, Sean Carroll, and Lauren Ouellette (ACO, 2014). (Aaronson and Carroll are physicists at MIT and Caltech, respectively; Ouellette was an MIT student.) They present and defend a formal definition of complexity – and while they never address this in their paper, I think it makes sense for us philosophers to treat complexity as the opposite of simplicity.

Given the past hypothesis, the standard view from physics is that the entropy of our universe increases pretty much monotonically – from a low-entropy state at the beginning of the universe to a high-entropy state at the end. (Picture time on the x-axis and entropy of the y-axis – other than random rare fluctuations, the slope of the line is always flat or positive.) The ACO measure of complexity for the universe, on the other hand, is like an upside-down U – the universe is simple at the initial low-entropy state, and simple at the final high-entropy state, while it is complex in the middle.

ACO write that, before their paper, there was

no general principle that quantifies and explains the existence of high-complexity states at intermediate times in closed systems. It is the aim of this work to explore such a principle… (2014, 2)

Their paper makes plausible that the universe would have high-complexity states at intermediate times – the time of our existence being such an intermediate time, and the current state of Earth being an example of a high-complexity state. They do not attempt to provide a justification for why the universe started in a low-entropy state – but perhaps here is where philosophical considerations can come into play. Given an account of simplicity that is well-grounded in physics, philosophers can then argue that simple states are likely to exist as the boundary conditions for our universe – that the initial and final conditions of the universe would be simple ones. So this is my tentative proposal for a naturalistic alternative to the theistic hypothesis that avoids the Boltzmann brain problem and saves induction: postulate the past hypothesis, and motivate it not only by pointing out that it saves induction, and the reliability of memory, but also that it is simple, and this gives foundational metaphysical reason to endorse it.

 Granted, the naturalistic past hypothesis is implausible – because physics tells us that a low-entropy state for the universe is highly improbable. But by pointing out that the past hypothesis postulates a simple state, and that (perhaps) simplicity is a guide to truth, we can provide some argument that makes the naturalistic past hypothesis more plausible. In doing so, we can make the naturalistic past hypothesis a more worthy competitor to the theistic hypothesis, with regard to the attempt to account for why our inductive practices will continue to be successful. It’s not a great argument but it’s the best I have to offer.

This main event doesn’t end with a knockout. We’re left to judge: which hypothesis is less implausible, the hypothesis that you are a Boltzmann brain, or the naturalistic past hypothesis, or the theistic hypothesis? That will have to be the reader’s decision.

**5. Epilogue: Self-Undermining Arguments**

I had coffee with Lily the other day. “Bradley”, she said, “I just found out that I have this strange and rare disease. The main symptom is that my memory of phone conversations is highly unreliable.” “Wow”, I replied. “How did you find out about this?” “Well”, she said, “my doctor called me the other day and told me so.”

There’s something funny going on in the above story. The funniness can be brought out if we try to convert that conversation into an inductively strong argument:

*Premise 1*: Doctors reliably tell their patients the truth about what diseases the patients have.

*Premise 2*: Lily reliably tells me the truth about her diseases.

*Premise 3*: Lily reliably remembers important phone conversations.

*Conclusion*: Lily told me the truth about her phone conversation; she does have this disease.

The problem is that, once we accept the truth of the conclusion, we have reason to doubt the truth of one of the premises – we have reason to doubt that Lily’s memory of important phone conversations is reliable. But once we reject that premise, we have no reason to accept the truth of the conclusion – the argument is no longer inductively strong. But if we don’t accept the truth of the conclusion, then we no longer have any reason to doubt that premise. And if we have no reason to doubt that premise, then the argument is once again inductively strong. And so on…

This is an epistemologically vexing situation, and unfortunately I don’t know of any epistemological literature to which we could turn to definitively resolve it (though for some discussion see Hume (1739, I, IV, i), Plantinga (1993, 234-5), and Talbott (2002, 157)).

It would be one thing if the vexingness just applied to this artificial Lily example, but it applies to a wide variety of cases. Consider, for example, Michael Huemer’s (2005) argument for the claim that when one is not an expert on a topic, one shouldn’t engage in critical thinking to decide what to believe; instead one is better off believing whatever the majority of the experts believe. If a non-philosophy-expert reads Huemer’s paper and, using her own critical thinking skills, comes to believe the conclusion, she might then decide to survey philosophy experts about Huemer’s conclusion. But most philosophy experts reject Huemer’s conclusion, so she would be led to reject it too. But once she rejected it, she’d be back to using her critical thinking skills, and those lead her to accept the conclusion of Huemer’s argument, and the circle continues.

The reason I bring all this up is that a key move in the previous section of the paper relies on a self-undermining argument. You have an understanding of the laws of physics, and based on that understanding, you get the result that you are probably a Boltzmann brain. But Boltzmann brains would not be expected to have true beliefs about the physical world – there is no reason to expect a rare random fluctuation out of a high-entropy state to produce a brain that has an understanding of the laws of physics of the world it’s appeared in. So if you run through this argument and reach the conclusion that you are probably a Boltzmann brain, you will recognize – now thinking that you are probably a Boltzmann brain – that you have no reason to believe what you thought you initially understood about the laws of physics. But then, you have no reason to believe that you are a Boltzmann brain. But then the circle continues – once you reject that hypothesis that you are a Boltzmann brain, you go back to thinking that you have an understanding of the laws of physics, and based on that understanding, you get the result that you are a Boltzmann brain … and so on.

So how should we handle this? Plantinga (1993, 235) briefly suggests that one should recognize that a situation like this is untenable, and hence one should not believe the initial premise. Plantinga admits that the premise might be true, but he says that it is irrational to believe it. His discussion takes place in the context of his evolutionary argument against naturalism, and he thinks that the key premise one should give up is naturalism. What is the premise that one should give up in Boltzmann brain situation? Plantinga might be happy to say “again, naturalism”. But naturalists would typically say that we should give up the premise that initial condition of the universe is a high-entropy state, and instead endorse the past hypothesis.

What worries me regarding both these moves (giving up naturalism, or giving up the premise that the universe starts in a high-entropy state) is that there is no evidence to merit giving up either of those premises. At least, there is no evidence to merit giving them up within the confines of the topic under discussion, the reliability of our inductive practices. Many theists would say that there are other reasons to believe theism (such as religious experience, the ontological argument, or what have you). But are there other, naturalistic, reasons that we should believe something like the past hypothesis, besides that it gets rid of the undesired result that one is probably a Boltzmann brain whose memories are unreliable and whose inductive beliefs about the future are mostly false? I have hope that there are such reasons, based perhaps on the theoretical considerations I raised above regarding simplicity. But it remains to be seen how much value there is in that hope.

**References**

Aaronson, Scott, Sean Carroll and Lauren Ouellette (2014), “Quantifying the Rise and Fall of Complexity in Closed Systems: The Coffee Automaton”, arXiv:1405.6903.

Albert, David (2000), *Time and Chance*, Harvard University Press.

Callender, Craig (2004), “Measures, Explanations and the Past: Should ‘Special’ Initial Conditions be Explained?”, *British Journal for the Philosophy of Science* 55: 195-217.

Cowling, Sam (2013), “Ideological Parsimony”, *Synthese* 190: 3889–3908.

Dawkins, Richard (2008), *The God Delusion*, Mariner Books.

Earman, John (2006), “The ‘Past Hypothesis’: Not Even False”, *Studies in History and Philosophy of Modern Physics* 37: 399-430.

Lewis, David (1986), *On the Plurality of Worlds*, Blackwell.

Huemer, Michael (2005), “Is Critical Thinking Epistemically Responsible?”, *Metaphilosophy* 36: 522-531.

Hume, David (1739), *A Treatise of Human Nature*, London: John Noon.

Plantinga, Alvin (1993), *Warrant and Proper Function*, Oxford University Press.

Price, Huw (2002), “Boltzmann’s Time Bomb”, *British Journal for the Philosophy of Science* 53: 83-119.

Swinburne, Richard (1979), *The Existence of God*, Oxford University Press.

Talbott, William (2002), “The Illusion of Defeat”, in James Beilby, ed., *Naturalism Defeated?* (Ithaca: Cornell University Press), pp. 153-164.