August 1, 2020

You are not a Boltzmann Brain[1]

John D. Norton[2]
Department of History and Philosophy of Science
University of Pittsburgh
http://www.pitt.edu/~jdnorton

According to the "Boltzmann brain" hypothesis, we popped into existence as a thermal fluctuation in an otherwise chaotic universe, with our brains replete with spurious memories of a fictitious, orderly past. The hypothesis extends less ambitious argumentation by Ludwig Boltzmann in the late 19th century, but it lacks the physical foundation of Boltzmann's original arguments. We are assured of neither the recurrence nor the reversibility of the time developments of the applicable physics. The Boltzmann brain scenario is much more likely to produce a physically spurious "batty brain" whose memories fail to conform to the scientifically well-behaved regularities of our brains.

1. The Earlier Proposal

It all started in the later nineteenth century when Ludwig Boltzmann needed to show that the time reversibility of micro-physics was compatible with the unidirectional increase of thermodynamic entropy of ordinary thermodynamics. His solution was ingenious. Imagine that the universe has come to thermal equilibrium. Since it is molecular in constitution, the equilibrium is dynamic and the state bounces around its mean state. A gas for example is, on

---

[1] This paper first appeared as "Goodies" page in Jun. 2015 at http://www.pitt.edu/~jdnorton/Goodies

[2] The ideas in this note developed in discussion with James Norton, then at the Department of Philosophy, University of Sydney. We also thank Wayne Myrvold for helpful discussion.

average, uniformly distributed in its accessible space. However random molecular motions will lead it to be momentarily slightly more dense here and slightly less dense there. Larger fluctuations are possible. A very big fluctuation might lead to the gas momentarily compressing to one side of a chamber. We would then find the gas momentarily in a low entropy state prior to its spontaneous, chamber-filling expansion.

The probabilities of such a spontaneous compressions are small. They are not just small, but astonishing, amazingly small. An ordinary volume of gas might have 10^{24} molecules moving independently of each other. The probability that all of them just happen to be in the left half of the chamber is (1/2) raised to the power of 10^{24}. That comes out to be something like 0.000--about 24 zeroes--01.

Being vastly improbable does not make something impossible. So, imagine a simple universe that contains just a gas in a chamber. If the gas is compressed to one side of the chamber in a low entropy state, it could have gotten that way by a spontaneous fluctuation from an equilibrium state. Or it could have evolved from a still lower entropy state. Perhaps it was at some earlier time compressed to one fourth the volume by a different spontaneous fluctuation. The first possibility is quite improbable. The second is even more improbable, for it requires the spontaneous appearance of an antecedent state of still lower entropy. In this scenario, the low entropy state most likely arose by fluctuation from a higher entropy state.

With that conclusion, Boltzmann had found a way to reconcile the time reversibility of micro-physics and the unidirectional increase of thermodynamic entropy. Pick some low entropy state. It is most likely to have arisen from a high entropy state by a fluctuation and it will most likely to return to a high entropy state. That is, entropy increases from that low entropy state both into the past and into the future. Contrary to what you might imagine, that thermodynamic entropy increases with time does not give us a preferred direction in time after all.

2. Boltzmann Brains

Boltzmann's argument has the look of something quite robust. It was laid out here for a kinetic gas, the subject of much of Boltzmann's investigations. Why not apply the argument to us as sentient observers in the world? We are low entropy systems. How do we know that we are not momentary fluctuations from some higher entropy state? The gas fluctuated to a compressed state that we would ordinarily assume had to come from a still more compressed state in its past. Why not say the same of us? We pop into existence as thermal fluctuations with our brains full of memories of a nonexistent past. That is the proposal of Boltzmann brains, which first appeared under that name in the work of Andreas Albrecht and Lorenzo Sorbo.[3]

3. Cartesian Deceiving Demons

It is an intriguing proposal. If it is right, then all we think we know is an illusion. The distant past of dinosaurs is just a false planetary memory, written deceptively in fossils created by fluctuations. So also is the whole story of the universe coming to be out a big bang, whose detritus collapsed to form stars and the planet on which we stand. It is not just the distant past. In the Boltzmann brain scenario, you most likely came into being by fluctuation in this instant. So your recollection of reading the sentence before this one is just as fabricated as every other memory.

That all that is needed is the momentary existence of your deceived brain means that the proposal minimally requires only one brain momentarily to exist. If that extreme solipsism is too hard to swallow, one could also use the same argument to establish that fluctuations can produce a sizable portion of the universe, complete with multiple sentient brains all agreeing on a fabricated past. (We shall see below that one of the more careful arguments for the Boltzmann brain scenario requires that it is this larger portion of the universe that appears via fluctuations.)

Wayne Myrvold has made what I think should be the first and, quite possibly, the last response.[4] This Boltzmann brain story requires us to think that all our past experience is illusory, just as does the original Cartesian deceiving demon. We think that the accumulated experience of scientists gives us good evidence to think that we emerged from a big bang some

---

[3] Andreas Albrecht, "Cosmic Inflation and the Arrow of Time," in *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity*. J. D. Barrow, P.C.W. Davies, & C.L. Harper eds. Cambridge University Press (2004). http://xxx.lanl.gov/abs/astro-ph/0210527 Andreas Albrecht and Lorenzo Sorbo (September 2004). "Can the universe afford inflation?" *Physical Review D* **70** (2004) 063528 http://arxiv.org/abs/hep-th/0405270

[4]Wayne C. Myrvold, "Probabilities in Statistical Mechanics," p. 17. http://philsci-archive.pitt.edu/11019/

10^{10} years ago. Instead we are to believe that we are in a universe at thermal equilibrium which, over some unimaginably vast period of time, finally manages to spit out a brain just like ours, as an extremely unlikely thermal fluctuation, complete with false memories of a still lower entropy past.

That requirement is self-defeating. For it also undermines all our knowledge of thermal physics. All the experimental evidence for all of science is a fabricated memory. We may also just be momentarily deluded into accepting the cogency of all the reasoning that connects these fictional experiments with the main result of science. In this skeptical scenario, we have no reason to believe the statistical physics that would generate Boltzmann brains in the first place.

4. Is the Argument Viable?

It is tempting to leave things at that. When Descartes imagined a deceiving demon who somehow fabricated all his experiences, he had no credible science to underwrite how this deception could be perpetrated. Has that now changed? Do we now have a scientifically viable way of producing a Cartesian deceiving demon, even though it proves to be self-defeating? We do not need to accept the viability of the Boltzmann brain argument. Here I will argue that the science itself requires so many speculative steps as to undo it. I will also suggest that there is a further problem in the argument.

5. Does the Physics Really Work?

Boltzmann's original argument tells us that a thermal system at equilibrium will eventually fluctuate away from equilibrium and that very large fluctuations are possible, even though of hugely small probability. That much was sufficient for his purposes. To demonstrate the compatibility of ordinary thermodynamics and a time reversible physics all he needed was low entropy states arising as fluctuations, with higher entropy pasts and futures.

The Boltzmann brain argument takes an extra step. It is, perhaps, a step that Boltzmann might eventually have had to take himself. But I don't believe he did. The extra step goes beyond fluctuations from thermal equilibrium that produce a low entropy state. It is that the low entropy state has to be one just like the state we are in now. It might be the momentary existence of a brain just like ours. Or it might be a larger piece of the universe that looks just like ours, complete with brain-embodied sentient observers.

That a *particular* low entropy state like this can be produced by a fluctuation requires more than the existence of a fluctuating thermal state of equilibrium. We need further arguments to establish that this particular sort of state can arise by thermal fluctuations. I can see two arguments for the extra step.

1. The recurrence argument

A familiar result of statistical mechanics is the Poincaré recurrence theorem. It tells us that the time evolution of a thermal system will eventually bring the system back to within an arbitrarily small neighborhood of its starting point. Hence, if ever it was possible to have a world with observers such as ourselves, then, even if we degrade thermally, over time we will eventually be reconstituted near enough; and it will happen infinitely often over an infinity of time. This argument proceeds from the fact that brains like ours can arise, we believe, through ordinary evolutionary processes. Since they arise through these processes only in the context of larger hospitable environments, the recurrence argument can only give us recurrence of these larger scenarios.

This recurrence argument cannot give us isolated brains that exist momentarily, since we would have to assume their possibility in the first place. That would make the argument circular.

2. The time reversal argument.

We believe that an isolated brain or a world with sentient observers such as ourselves will, over time, degrade thermally into a final, equilibrium heat death. If our micro-physics is time reversible, then the reverse process can happen: a state of thermal equilibrium can evolve in the reverse direction into an isolated brain or a world with sentient observers such as ourselves. They would appear as a highly improbable fluctuation from thermal equilibrium.

Once we lay out these arguments explicitly, it is easy to see that they are quite fragile. Here are some of the problems.

6. The Recurrence Problem

The difficulty with the recurrence argument is that it relies too heavily on an analogy with the kinetic theory of gases (which was the case dealt with primarily by Boltzmann). The Poincaré recurrence theorem obtains for kinetic gases confined in a vessel. Such a gas will explore its state space thoroughly over eons of time. It is tempting to imagine that this theorem will do the work for us with brains as well. However, there is no assurance that it will and reasons to think that it will not.

The Poincaré recurrence theorem has antecedent conditions that must be met and are likely not met. Two important ones are:

> **Finite phase space volume.** Phase space is the product space of spatial and momentum degrees of freedom. It is finite for a gas confined to a box. If that finitude fails, recurrence fails. For example, a gas expanding into an infinite space can continue expanding indefinitely. The recurrence theorem cannot assure us of a return.
> Our space is quite possibly not finite, depending on which cosmology is correct.
>
> **Finitely many degrees of freedom.** That is, the phase space has finitely many coordinates. In the case of a monatomic gas, there are 6 coordinates for each atom: 3 for spatial position and 3 for the direction of motion. Since there are finitely many atoms, the total count of coordinates is also finite, if large.
> Brains and sentient observers consist of vastly complicated biochemical systems; and they would be a part of a potentially infinite cosmology. Are we assured that this total system, brain and cosmos, has only finitely many degrees of freedom?

And finally...

> **Quantum concerns.** Brains are not, in the end, classical systems; and our universe is also quantum mechanical. Poincaré's theorem applies to classical systems. Is there an appropriately strong quantum analog for a brain bearing cosmos?[5]

7. The Turning Point Problem

There is a second problem associated with the analogy to the kinetic theory of gases. It is easy to imagine a gas spontaneously recompressing through a fluctuation, pausing at the moment of lowest entropy and then expanding back to a higher entropy state. This momentary

[5] For some quantum versions of the theorem, see David Wallace, “Recurrence Theorems: a Unified Account,” http://philsci-archive.pitt.edu/9838/ The result cited as the “quantum recurrence theorem” assumes a discrete energy spectrum for the Hamiltonian, which is too restrictive for the general case needed here.

turning point just involves molecular trajectories all passing through a momentary region of high molecular density.

Are we assured that a similar turning point is possible for sentient brains? The (presumed) time reversibility of our micro-physics assures us that our processes could run in reverse: we would grow younger instead of aging and our memories of events would empty as the events are undone. The Boltzmann brain scenario requires more. It requires that this reversed process can come to a turning point, where the reversed aging halts and the system turns around and proceeds with normal aging.

The time reversibility of our processes does not assure us that such turning points are possible. That they are possible is an extra assumption for which further argumentation is needed.

There is an analogous problem at the initiation of the fluctuation in the state of thermal equilibrium. Call it the "*initiation problem.*" While the time reversed reconstitution of brains is possible, without a Poincaré recurrence theorem, what assurance is that this particular fluctuation will be initiated, even if only in probability?

8. Time Reversibility Problem

It is assumed that the micro-physics of brains and worlds is time reversible. That is not so assured. The main problem is that these systems are quantum mechanical.

First, the weak interaction in particle physics is not time reversal invariant. One might wonder whether a mere "weak" interaction can matter. Can we not just file it away as too weak to matter? We cannot do that. Time reversibility arguments require a time reversible physics. If that reversibility fails, the argument cannot be mounted. Analogously, we cannot argue in all generality against the possibility of a perpetual motion machine if the law of conservation of energy has exceptions, even if weak.

As to whether the weak interaction affects biological systems, recall that the weak interaction mediates in radioactive decay. Radioactivity can cause genetic mutations and these mutations can figure essentially in a particular evolutionary history. Thus we cannot be assured that such an evolution history can run in reverse. Its physics is not time reversible.

Second, the standard account of quantum mechanical measurement involves a time-irreversible collapse of a quantum wave on measurement. If that collapse is a real process, then the micro-physics is very far from time reversal invariant. Quantum measurement remains a hotly debated topic in philosophy of physics. However, to presume time reversibility of the quantum processes of life is to presume a particular resolution of the problem that somehow protects time reversibility.

9. Boltzmann Brains are less Probable than Batty Brains

There is a second general problem whose import is less clear to me. I will rehearse it here and leave the reader to decide the import.

For the Boltzmann brain skeptical scenario to succeed, it must show that, when fluctuations produce brains, they are brains just like ours. The scenario seems to fail this test. For a natural reading of its arguments indicates that brains just like ours would be extremely improbable among the overwhelmingly more common "batty brains" produced by the fluctuation process.

To develop the problem let us grant that brains replete with false memories can be created by fluctuations. Granting that does not grant that the brains produced will be just like ours. All sorts of variations are possible. Our memories are of a relatively orderly past, full of regular occurrences conforming to discoverable laws. Nothing forces a Boltzmann brain to have such regular memories. Since they are entirely fabricated, there could be any wild departure from regularity.

For example, we recall that the sun rose every day for the past 365 days of the last year. In fabricated memories, the sun may or may not rise on each of those days. There's only one past in which the sun rose every day. There are enormously many more, 2^{365}, in which they rose on some days and some days not.

That might seem to settle the matter: it is vastly improbable that a Boltzmann brain might have memories of a regular past just like ours. However, there is a loophole: the brain does not have to have memories of a regular past just like ours. It just needs to think, even if only for a moment, that its memories are of a regular past. That belief may be mistaken, but all it needs to have is just that belief.

This loophole is harder to close, but its effect can be greatly diminished. Assuming that our appraisals of cogency are all illusions reinforces the original concern that the Boltzmann brain argument is self-defeating. Further, since all that is at issue is a belief, there are correspondingly many possible alternatives for this belief state. It may be one in which we believe our memories of sunrises every day over the past year are veridical. But it may also be one in which we recognize that many of our memories are entirely spurious; and this could happen in many possible ways. There are many combinations of veridical and spurious for our confidence in our memory of the past year's sunrises. It is the same combinatorial factor of 2^{365}.

However we approach this problem, it is clear there are many possible belief states. For the Boltzmann brain scenario to succeed, it has to assure us that a fluctuation is likely to produce a brain that instantiated just one: the one that is convinced of the veracity of its memories of sunrises.

Note that the veracity of memory at issue is not of the recall of things where we know that our memories are fallible. We glimpse the face of a robber and then try to pick out the perpetrator from a line up. Research in psychology has shown such recollections can be quite fallible. What is at issue is the memory of occurrences that are quite reliable. We firmly recall that the last 365 days were all quite ordinary in that the sun rose and set as it always did. Nothing in the Boltzmann brain scenario assures us that a sentient brain must have such uniformity in its beliefs about the veracity of these mundane memories.

In short, what is most likely under the Boltzmann brain scenario is that brains are batty brains, unlike ours. They have chaotic memories of irregular pasts and equally chaotic beliefs about the veracity of ordinary memories.