Types of Boltzmann Brains

*Alexey Turchin*

Digital Immortality Now

Foundation Science for Life Extension

alexeiturchin@gmail.com

*Roman Yampolskiy*

University of Louisville

**Abstract**. Boltzmann brains (BBs) are minds which randomly appear as a result of thermodynamic or quantum fluctuations. In this article, the question of if we are BBs, and the observational consequences if so, is explored. To address this problem, a typology of BBs is created, and the evidence is compared with the Simulation Argument. Based on this comparison, we conclude that while the existence of a “normal” BB is either unlikely or irrelevant, BBs with some ordering may have observable consequences. There are two types of such ordered BBs: Boltzmannian typewriters (including Boltzmannian simulations), and chains of observer moments. Notably, the existence or non-existence of BBs may have practical applications for measuring the size of the universe, achieving immortality, or even manipulating the observed probability of events.

**Disclaimer and trigger warning:** some people have emotional breakdowns when thinking about topics described in the article, especially the “flux universe”. However, everything eventually adds up to normality.

[1. Introduction 2](#_Toc12278205)

[2. Possible types of BBs 3](#_Toc12278206)

[2.1. Biological brains vs. observer moments 3](#_Toc12278207)

[2.2. Physical types 3](#_Toc12278208)

[2.2.1. Mathematical BB 3](#_Toc12278209)

[2.2.2. Thermodynamic BBs, or “rock minds” 4](#_Toc12278210)

[2.2.3. Quantum fluctuations create minds 5](#_Toc12278211)

[2.2.4. De Sitter vacuum nucleation 5](#_Toc12278212)

[2.2.5. Qualia BBs 6](#_Toc12278213)

[2.2.6. Modal realism also implies BBs 6](#_Toc12278214)

[2.2.7. Everettian BBs 7](#_Toc12278215)

[2.2.8. Artificial BBs 7](#_Toc12278216)

[2.2.9. Panpsychism 7](#_Toc12278217)

[2.3. Boltzmann typewriters 7](#_Toc12278218)

[2.4. Observer moments and BBs may have convolutional structure 8](#_Toc12278219)

[3. Possible observational consequences of being BB 9](#_Toc12278220)

[3.1. Why the statement that “I can’t be a BB as my experience is not random” is wrong 9](#_Toc12278221)

[3.2. Pure BB-world made of chains of observer moments 9](#_Toc12278222)

[3.3. Miracles, magic, viruses and glitches in chains of observer moments 10](#_Toc12278223)

[4. Measuring the most probable types of BBs 10](#_Toc12278224)

[4.1. Analogy between simulation argument and BBs 10](#_Toc12278225)

[4.2. Measure of BBs 11](#_Toc12278226)

[4.3. Extremely simple or extremely complex BBs should dominate 11](#_Toc12278227)

[4.4. A map of possible observers and BBs in the universe 12](#_Toc12278228)

[4.5. Escaping the BB curse: BBs as reflections of real observers 13](#_Toc12278229)

[4.6. Stable attractors for chains of observer moments 14](#_Toc12278230)

[4.7. Superposition of interpretations 14](#_Toc12278231)

[4.8. Non-locality of the mind in the multiverse as a way to escape BBs 15](#_Toc12278232)

[4.9. Biological evolution as a shadow of my existence 16](#_Toc12278233)

[5. Practical application of BBs 16](#_Toc12278234)

[5.1. No BBs means a “Big Rip” soon, and astronomical waste for future supercivilizations 16](#_Toc12278235)

[5.2. Boltzmannian immortality 17](#_Toc12278236)

[5.3. “Permutation City” immortality 18](#_Toc12278237)

[5.4. Exploiting glitches in the chains of observer moments 18](#_Toc12278238)

[5.5. Flux universe 18](#_Toc12278239)

[5.6. BB uncertainty and the problem of constructing of a global prior 19](#_Toc12278240)

[Conclusion 19](#_Toc12278241)

[Literature 19](#_Toc12278242)

I close my eyes only for a moment, and the moment's gone

All my dreams pass before my eyes, a curiosity

Dust in the wind

All they are is dust in the wind

*Kansas*

# 1. Introduction

The idea of freak observers, called Boltzmann brains (BB), became popular in the beginning of the 21st century with the works of (Dyson, Kleban, & Susskind, 2002) and (Albrecht & Sorbo, 2004). BBs are observers’ minds which appear not as a result of natural evolutionary processes but as a result of some random stochastic process. This idea raises several questions:

1. Are freak observers the dominant type of observer in the universe?
2. If yes, how I can prove to myself that I am not such an observer, and more generally, what are the observational consequences of being a BB?
3. If no, what does this tell us about the size and structure of the universe?

Recently, Carroll (Carroll, 2017) argued that typical objection against BBs, that our observations are not random is insufficient and that the whole theory of BBs should be rejected as it is “cognitively unstable”. Such “cognitive instability” is—according to my partly stable understanding—that if we are BBs, we can’t create true theories, and thus if a theory predicts that we can’t create correct theories, it is “cognitively unstable”. In short, it could be compressed into a conjecture: “If statement X is true then the truth of X can’t be measured”, similar to the liar’s paradox (self-refuting statements like "I am lying"). However, cognitive instability is not evidence that we are not BBs. Don Page wrote that invoking a new scientific principle just to escape BBs is non-economical, and instead uses Bayes’ theorem to disprove BBs (Page, 2017). However, Page’s criticism is based on our ability to correctly distinguish that our observation is ordered, which is not granted for BBs, as we will discuss in Section 3.

De Simone (De Simone et al., 2010) suggested in a section “Boltzmann Brain Nucleation Rate” a classification of BBs starting from “Solar system BB” and to smaller and smaller ones, including three main types: “a human-like brain, a near-black hole computer, and a diffuse computer”. A diffuse computer is a clod of gas which is somehow is capable of producing computations. They also introduce the idea of a “minimal BB”, which however, is not, in their interpretation, a single observer-moment, but a functional analogue of human brain. They add: “Thus, the formula suggests that diffuse-gas-cloud Boltzmann brains of radius 200 light-years can be thermally produced in our universe, at the present time, without suppression! If this estimate were valid, then Boltzmann brains would almost certainly dominate the universe”, but later conclude that it can’t happen as it would require unusually cold regions.

To solve these problems, we first will create a typology of possible BBs, then explore their observational consequences and logical puzzles connected with the idea. We will then address practical application of BBs, like Boltzmannian and “Permutation City” immortality.

# 2. Possible types of BBs

## 2.1. Biological brains vs. observer moments

There are two possible extreme types of BBs: full biological brains and observer-moments (OM), which are single moments of subjective experience. A full-brain BB requires that around 1026 atoms collect in the exact order that represents an observer’s brain, which is a very low probability event, even under “favorable” conditions in the primordial soup of amino acids.

For an observer-moment BB to arise, only computations which represent the observer-moment (OM) are necessary. The typical OM in humans is around 106 bits, as measured by the size of the visual field, and has a duration of around 0.1 second. In that time, only around 10 subsequent neurons will be able to fire, so the computation in a neural net which produces the human OM is relatively simple. (A lot of complex processes in the brain are too slow to be presented inside an observer moment, so we should look only at neuron firings and synaptic weight, but not at slow gene expression.) Very roughly, we can estimate that 10 billion neurons need to fire 10 times to create an OM, and each neuron has 1000 synapses, so there are ~1014 synaptic calculations per OM. This is probably an upper limit, as most such calculations are redundant or unobservable, and the real number of computations needed to indistinguishably mimic an OM is likely closer to 106.

A biological BB could exist for more than one “tick” of time and create several OMs, but each random OM, which is emulated by some non-biological computational process, is very short, and ends completely with that OM. However, a biological BBs with a few ticks could have time to understand its situation and reason more logically, like: “OMG, I am a naked brain in a river, I will die soon!” or “I am in empty space and I am freezing…” See also fictional [internal monologue](https://www.goodreads.com/quotes/198068-another-thing-that-got-forgotten-was-the-fact-that-against) of a whale randomly appeared in space in Douglas Adam’s “Hitchhiker’s guide to the galaxy”.

Only biological BBs (or close-to-such, functional analogues of a brain) could have coherent thoughts, if they are capable to exist more than just one observer-moment. However, such BBs are more complex and thus many orders of magnitude less probable than one observer moment long BB (OM-BBs).

## 2.2. Physical types

### 2.2.1. Mathematical BB

If a pure Tergmarkian mathematical universe exists (Tegmark, 2014), it should include all possible Turing computers, and if an observer mind could be successfully simulated in a computer (no hard problem of consciousness), it would include all possible observers. Any mind-state could also be represented (again, if no qualia and causality are required) as a string of numbers. Note that the class of all possible Turing computers include not only observers, but also advanced AIs of high complexity, and the number of different possible AIs will be even higher than the number of pure observers. Such AIs can run simulations of whole worlds, which recursively include simpler observers. This may affect the global measure of observers in a mathematical world, as it will compensate for the otherwise-expected domination of completely random observers.

So, there are three main types of mathematical BBs:

1. *Strings of numbers*, equal to observer moments. Mueller showed that even the existence of such strings is sufficient, as it is possible to create a complexity-based function which transforms one string into another, and the result would eventually look like the physical world (Mueller, 2017). However, if we assume the existence of number strings, why not assume the existence of more complex mathematical objects, including computers?
2. *Possible calculations*,which represent just one observer moment. This is relevant if we assume that an act of experience is some form of calculation, for example, an act of image recognition of, say, a cat. As an act of experience is – in first approximation – an act of image recognition is a neural net inference calculation.
3. *Possible Turing computers*, simulating any possible observer.
4. *Possible AIs*, which are simulating observers.

### 2.2.2. Thermodynamic BBs, or “rock minds”

Thermodynamic BBs are those which are created via classical stochastic processes, like atoms aggregating to create a brain. As any brain can decay, and all low-level laws of physics are time-invariant, then—albeit with extremely low probability—a brain could coalesce from scattered atoms, which, based on Poincare’s recurrence theorem (Furstenberg, 1981), will eventually happen in any closed system. Such a brain could appear in the primordial ocean of some planet that is otherwise empty of life. Another example of such a brain could appear in an almost empty universe if particles happen to collide in just the right way. Alternatively, an observer moment analogue could appear if the same calculations or causal structure of the OM is produced by some random process. This example is close to that Egan described in his dust theory (Egan, 2010).

A few types of possible thermodynamic BBs could be suggested:

1. *Biological minds on planets*, appearing directly from a primordial soup. Even if some reptile is intelligent enough to become a conscious observer following some series of random mutations, this example could still be considered a random biological mind.
2. *Biological minds appear in pure stochastic universe,* as a result of the matter clumping. Here, some atoms collide and form the brain in an otherwise empty post-heat-death universe.
3. *Observer moments which appear from dust (stone minds) (Egan, 2009).* Some thermodynamic billiard could have the same causal structure as computations in a mind, creating an observer moment. These could also be called minds-in-stone. In other words, the random collision of atoms in a gas (or thermal movements of atoms in a solid body) may form a causal structure, similar to the causal structure of an act of subjective experience. Whether or not this is possible is unclear, as we still don’t know how subjective experiences arise.

Putnam and Chalmers have discussed a possibility of computations inside natural physical processes, as detailed in “Does a Rock Implement Every Finite-State Automaton?” (Chalmers, 1996). In the abstract Chalmers states:

Putnam has argued that computational functionalism cannot serve as a foundation for

the study of the mind, as every ordinary open physical system implements every finite-state

automaton. I argue that Putnam’s argument fails, but that it points out the need for a

better understanding of the bridge between the theory of computation and the theory of

physical systems: the relation of implementation. It also raises questions about the classes

of automata that can serve as a basis for understanding the mind. I develop an account of

implementation, linked to an appropriate class of automata, such that the requirement that

a system implement a given automaton places a very strong constraint on the system.

More generally, any complex physical process could be seen as a mathematical function, mapping initial states to outcomes, and thus performing complex computations, which could be interpreted as solutions of some problems. This is known as a “waterfall argument”—for more on this, see the analysis and criticism in this [paper](https://www.scottaaronson.com/papers/philos.pdf) by Scott Aaronson (Aaronson, 2012).

### 2.2.3. Quantum fluctuations create minds

A more modern idea, which has fueled interest in BBs, is that observers could appear from quantum fluctuations in de Sitter space. Heisenberg’s uncertainty principle states that the existence of such a fluctuation should be extremely short. If it is too short, there will be no time for any acts of observation, as a normal brain needs at least one tick of its internal clock to think a thought (around 0.1 second). But in the case of some functional model of the brain as BB, the objective existence time could be compressed relative to what the observer thinks about his-her subjective time of observation.

In any case, the need to move from an internal state 1 to an internal state 2 to create conscious experience significantly limits the possibility of quantum BBs.

1. *A quantum fluctuation creates an observer moment* by mimicking its causal structure, without cells or other physical media.
2. *Functional model of a mind*, appearing from a quantum fluctuation in vacuum or thermodynamic fluctuation. The functional model is a simpler version of the whole brain, which uses less atoms, weighs less, and thus is more probable. For example, a superefficient computer which runs a human upload may weigh just a few milligrams, which makes its appearance, via fluctuation or random matter clumping, much more probable than the appearance from nowhere of a whole biological brain (and a body supporting it, at least temporary).
3. *Biological mind appears as a quantum fluctuation* in an otherwise empty universe for a very short time, just long enough to have one thought.
4. *A quantum fluctuation creates a computer* which runs a model of human mind, which has just one thought.

### 2.2.4. De Sitter vacuum nucleation

Armstrong [wrote](https://www.lesswrong.com/posts/ygELzNSAF5nzLXD7j/are-you-in-a-boltzmann-simulation):

If we understand [dark energy](https://en.wikipedia.org/wiki/Dark_energy) correctly, it [dark energy] will transform our universe into a [de Sitter](https://en.wikipedia.org/wiki/De_Sitter_universe) universe. In such a universe, the continuing expansion of the universe acts like the event horizon of a black hole, and sometimes, spontaneous objects will be created, similarly to [Hawking radiation](https://en.wikipedia.org/wiki/Hawking_radiation). Thus a de Sitter space can nucleate: spontaneously create objects. The probability of a given object of mass M=1kg being produced is [given as](https://arxiv.org/pdf/1008.0808.pdf) exp[−M×1069]. This number is much, much, much, much, ...., much, much, much, much smaller than the quantum fluctuation probability. But notice something interesting about it: it has no time component. Indeed, the objects produced by nucleation are actual objects: they endure. Think a [brain in a sealed jar](https://womenwriteaboutcomics.files.wordpress.com/2013/08/futurama-space-pilot-3000-3.jpg), floating through space.

Based on this, he suggested that sealed brains’ OMs will dominate over quantum fluctuation OMs, as, being less probable, they could create many more OMs, compensating for their lower probability. Such a nucleated brain could also be called a “Boltzmann astronaut” (BA), and Armstrong assumes that some such BAs could be equipped with a very effective source of energy, maybe even antimatter, which would allow it to work for a very long time.

However, Donald Hobson replied that 2 kg BB nucleation is exp[−1069] less likely than 1 kg BB nucleation, and assuming that the additional 1 kg is energy support, it should be able to provide energy for exp[−1069] seconds, which seems impossible, as it is an extremely long time. A possible replacement of a biological mind on a small computer brain model doesn’t change this calculation, as again computers without an energy source are much more likely.

Linde et al (De Simone et al., 2010) suggested several types of BBs which could nucleate, ordered from less probable to the most ubiquitous:

* Solar system size BBs,
* Biological brain BB,
* BB, powered by computations on the border of small black hole,
* Diffuse gas BBs, where smallest amount of particle on lowest possible temperature is mimicking causal structure of the process inside an observer mind.

### 2.2.5. Qualia BBs

Let's assume that qualia actually exist as part of consciousness, and moreover, they have independent oncological status, so they can't be derived from physics. In other words, assume that qualia are new type of mathematical objects, and each quale is an axiom. Thus, they should not be derived from anything else, but instead are self-born. For example, anything which creates qualia of red should have something red in itself. Thus, red doesn’t depend on anything except red.

1. *Combinations of qualia*. BB as a random combination of qualia.
2. *One quale,* which, in an illusionary manner, represents multitudes of qualia, and is equal to just one observer moment.
3. *Ordered platonic qualia.* In this case, we assume that there are qualia of ideas, and that such qualia are related to the lower-level qualia of experiences. For example, the quale of space controls the qualia of colors. This creates a hierarchy of possible colors (similar to a platonic hierarchy of ideas), and any given observer moment is a random subset of a such hierarchy. Order grows from the higher level of qualia to the lower level, in the same way in which if we were to create a universe with a random set of laws, the things inside this universe would be more ordered than its laws.

There is a form of panpsychism which claims that any causal physical process must have internal experience. Argonov wrote that consciousness may be a property of a single electron (Argonov, 2012), while Yudkowsky wrote about “[how the algorithm feels from inside](https://www.lesswrong.com/posts/yA4gF5KrboK2m2Xu7/how-an-algorithm-feels-from-inside)” (E. Yudkowsky, 2008a). Such qualia-based panpsychism is essentially a claim that the whole universe is full of BBs, which are capable of having qualia, but not of reasoning.

### 2.2.6. Modal realism also implies BBs

If everything possible exists, all possible random minds also exist.

### 2.2.7. Everettian BBs

If Everettian multiverse is possible, then within it, every piece of vacuum has a non-zero measure of creating any possible brain, and thus, we could say that it exists in some sense.

### 2.2.8. Artificial BBs

Almond (Almond, 2003) suggested *indirect mind uploading via quantum randomness generator* as an instrument for mass resurrection. Such a generator creates a long string of numbers, which is put into a computer memory and treated as an executable file. In some Everettian branches, any possible mind will appear with very small measure (as well as any possible AI—this is limited only by the length of the string of numbers).

Moreover, such minds could appear “randomly” in our computers, in the case of “soft” disk errors caused by cosmic rays. But most computers have protection against this phenomenon, like checksum of critical files. We now have billions of computers running millions of seconds, and in theory, each could become a BB after some combination of soft errors (but multiplying by billions here doesn’t changes the overall probability distribution, which is dominated by a *very* large number of necessary combinations.)

The main feature of a such a computer-BB is that it will not, in most cases, disappear after the first moment of existence, and, if internally stable, could exist long after the computer is rebooted. This increases the chance of Boltzmannian simulations.

Another type of BBs in a computer is “[suffering subroutines](https://reducing-suffering.org/what-are-suffering-subroutines/?fbclid=IwAR2y8eP71cCNF6YL5uICmEIV04n_W8XzjdkcCuAq8cKZ9PYlsMEZK2eXoBc)” (Tomasik, 2019), that is, a computation which has an internal experiences of suffering, which could arise while training of neural nets via reward. Any random quale could appear inside a neural net which is performing some other task, and it could “feel” nostalgia or any other experience for a short moment. As we don’t know yet was is the neural correlate of consciousness and qualia, we don’t know which types of computations are capable to generate temporal observers. Moreover, even if some non-Turing computations are needed for subjective experiences, like quantum computers (Penrose & Gardner, 2002), they again could be mimicked by some natural quantum processes.

### 2.2.9. Panpsychism

One of the interpretations of panpsychism is that any causal process has some rudimentary phenomenological consciousness, or qualia (see e.g. Argonov’s “One electron theory of consciousness” (Argonov, 2012)). This idea has been rejected on grounds similar to the logic used to reject Boltzmann brains: that if it were true, I, as an observer, am unlikely to be a human person, but instead more likely to be a random causal process in the world. As we will see later, this argument doesn’t hold, as we should count only those observers who are able to ask this question.

## 2.3. Boltzmann typewriters

In a classical example, a monkey randomly hitting a typewriter for extremely long time could write a famous novel (for example, “War and Peace”). If two things could appear randomly *with equal probability* from noting: either a book or a monkey with a typewriter who will later print many random books, then most random books will be printed by such monkey.

We will call the type of fluctuation capable of creating a lot of observer moments inside it a “Boltzmann typewriter”. Boltzmann typewriters (B-typewriters) seem to be important, as they may be capable of outperforming both real-world and classical BBs in their ability to create random minds; they also create less-random minds than pure fluctuations do, which could explain why we are not in a completely random world.

The most obvious example of a B-typewriter is our universe itself, as it supposedly was created from quantum fluctuation, and the process of evolution on Earth is using something akin to B-typewriters to search for most well-adapted organisms. Thus, it creates 100 billion people, and each has at least a billion observer moments, so the total number of human observer moments is around 1020. However, we actually need to count only BB-aware observer moments, as we will discuss later. Another type of B-typewriter, Boltzmann simulations, an AI which appears out of nothing and runs a simulation of many minds, was suggested by Armstrong (Armstrong, 2018).

The most interesting question is why B-typewriters are not most probable type of the observer-generating environments? For example, if a cloud of gas appears at the size of Solar system and it collapses into the Solar system, it would be enough to trigger the creation of the Earth and many observers, but we will not observe the stars. One answer may be that it is simpler to start eternal inflation than to create just a finite amount of perfectly composed gas, as it must be enriched with heavy elements from nearby supernovas to create the Earth.

The classification of B-typewriters overlaps with the physical mechanisms of BBs, as any such mechanism could also create a B-typewriter.

1. *Generators of random numbers* which themselves randomly appear
2. *Physical universes* optimized to create minds
3. *Small, one-planet universes,* optimized to create minds.
4. *Minds directly appearing from primordial soup of amino acid which supposedly was capable to create first living cell (let’s call it “amino-soup”)*, which appear without previous biological evolution, but on physically real planets. Such a planet is a B-typewriter.
5. *Boltzmannian simulations*, random AIs appearing as quantum fluctuations. They may be more probable than biological fluctuated minds, as they weigh less.
6. *Random minds in real computer simulations*. If, according to Bostrom’s Simulation arguments (Bostrom, 2003), there will be many simulations run by future AI, such simulations will simulate more minds than real minds that have existed, and some such minds will be completely random; in other words, for any given mind there will be a simulation which will simulate it, even for completely random and chaotic minds.
7. *Quantum resurrection machine* described in section 2.2.8.

B-typewriters work as cosmological inflation for experience: they dilute the randomness of experience, and thus experience is more ordered in them, but still less ordered than it is in the completely real universe. In the same way, cosmological inflation dilutes the concentration of magnetic monopoles is our world and makes them completely unobservable.

## 2.4. Observer moments and BBs may have convolutional structure

Typically, a BB is assumed to be a structure-less clump of atoms or bits, like a string of numbers. In the same way, an observer moment is also assumed to be an object without structure, like a screenshot or a 2D picture of visually observable things. However, based on the internal phenomenological observations of consciousness, the consciousness is summing up different things, and as a result looks like a conus with many empirical details on the bottom and thought, moods and attention on the top. The data flows from the bottom to the top, and attention is directed from the top to bottom. The “surface” of the bottom “is built” from basic experiences like qualia.

An analogue of an OM could appear from of thermodynamic billiards of random atoms in gas hitting each other and thus having the same causal relations as some convolutional network act of inference. Such billiard may have a similar, pyramidal structure as a convolutional network: many atoms hit each other initially and the whole pyramid or relations converges in the just a few atoms hit the last one. Such pyramidal OM structure results in a situation in which randomness is accumulated to the bottom and “order” is accumulated to the top. For example, if we have a neural network which recognizes either cats or dogs, any random input in it will generate “cat” or “dog” output. It could be interpreted as if it finds some “order” in any inputs, which is similarity to cats or dogs. Obviously, it is only illusion of order.

In other words, *any act of observation is an act of finding an order in a random data*, and thus it should not be surprising that even random BB will find some order in their observations.

If this convolutional structure of an OM is accurate, it will have observational consequences for BBs: random moments will not be like TV static. Any randomness on the bottom will be interpreted as some structure on the upper level of convolution. This is similar to typical human experience: there is a lot of low-level randomness in human experience.

Look at the room around you: Objects are located rather randomly relative to each other, and order appears only when the mind classifies them as tables, chairs etc. Order, or the illusion of order, is just a property of a convolution neural net, which interprets many inputs as one output.

The convolutional structure of an OM means that many different low-level OMs have the same high-level outcomes. Many different images of cats result in the same observation of “cat”. This makes *chains* of observer moments, a concept which will be discussed later, even more probable.

Qualia-based BBs also could be structured BBs, but in that case, some different theory of qualia is needed to explain how qualia are connected with each other.

# 3. Possible observational consequences of being BB

## 3.1. Why the statement that “I can’t be a BB as my experience is not random” is wrong

The main objection to the BB theory is that most BBs are completely random minds, so they should experience something like TV static, or at least a very chaotic environment. However, I am not a completely random BB: I am a mind—a member of the class of mind—which is thinking about BBs. This means that my experience is not completely random.

Also, if I am a BB, my thoughts are random. There is no causal connection between a BB’s experiences and its thoughts. Thus, a BB could have a random experience and still think that it is non-random. In that case, my thought that my experience is not random is not evidence that it is ordered and doesn’t affect conclusion that I may be a BB. “I can see my room, but why do I think that I actually see it? What is my evidence that “my room” isn’t just a random thought?”?

## 3.2. Pure BB-world made of chains of observer moments

Randomness is a property of the world, but not of separate observer moments. For any random observer moment there exists another observer moment that makes the first non-random. For example, if I see TV static, my next observer moment could be that I turn my head away from a TV set. In that case, the first observer moment is logically explained by the second.

The idea of “logical explanation” can be formalized through the theory of computational complexity: the logical explanation will be the simplest one (Mueller, 2017). Such logical connection via simplest explanation creates chains of observer moments, which are not necessarily connected with each other. Mueller shows that such chains tend to self-stabilize and may look like ordinary life, and even derives basic laws of physics from the idea of chains of random observer moments.

## 3.3. Miracles, magic, viruses and glitches in chains of observer moments

While a chain of observer moments may locally look like a an observation of seemingly stable outside world world, it does not have any outside stabilizing mechanism (real world), and some inconsistencies may eventually appear. Mueller wrote that minds from different chains could communicate but not prove to each other that they have a certain type of inconsistency. Such inconsistencies may look like a “Mandela effect”, that is, different memories of different people about the past that do not converge, or different glitch in the matrix (GITM)-type effects. (In other work I have shown that reported stories about GITM are unlikely be evidence of real simulation, as most simulations will be able to debug such glitches (Turchin & Yampolskiy, 2019) ).

A chain of observer moments may work as a “hostless simulation”, an illusionary world without a creator. As there is no creator, there are no instruments to control the stability of the illusion, or these instruments are also non-conscious processes, working only on average (like the inability to photograph a strange phenomenon, as has often been reported). Hostless simulation could also be affected by self-replicating entities like viruses, which may look like repeating strange phenomena with absurd or random elements. Also, as observer-moments are connected via similarity, there could be parasitic connection with observer moments that are not close to each other in illusionary space-time, and this could look like coincidence, magic, or telepathy.

In the real world there are no glitches, but BB-chains should have bugs from time to time. We could learn more about the type of our world we live in by observing possible inconsistences. The types of possible inconsistencies should differ between computer simulations and BB chains. In the first case, they are linear (like a wrong texture), and in the second case, they defy time-space (Mandela effect = past inconsistency), as in BB-chains only *now moment* is real, but past and future are just its most probable “shadows”. However, most claimed GITM could be best explained by hoaxes, memory problems and cognitive biases, and can’t be considered real evidence.

## 3.4. Ignoring BBs based on Updateless decision theory

According to the group of decision theories based on the idea of updatedness, a person should act in the way which will benefit most of observers in his-her situation (Eliezer Yudkowsky & Soares, 2017). For example, I should go to voting, because if everybody will vote, the good candidate will win, despite the fact that one my vote is not changing anything. Any choices made by OM-long BBs may be ignored as these choices will not have any consequences, as the BBs will disappear in the next moment. So, one has to act as he if he is real person no matter how many BB-copies of him exist, as such actions will actually benefit the real copies of him.

# 4. Measuring the most probable types of BBs

## 4.1. Analogy between simulation argument and BBs

Crawford mentioned the analogy between BB and simulation in “[Freak Observers and the Simulation Argument](https://onlinelibrary.wiley.com/doi/pdf/10.1111/rati.12009)” (Crawford, 2013). In both cases, the numerical abundance of non-conventional observers suggests that we should be such ourselves. This logic means that either both arguments are false, or both are correct. If both are true, what does it mean to be both a BB and in a simulation? This question could be addressed in two ways: either a whole simulation appears randomly as a BB, or some BB-like observers exist in the simulation of the real world.

## 4.2. Measure of BBs

Bostrom’s “Bland indifference principle” claims that the odds of being a biological brain or its computer copy are equal (Bostrom, 2003). However, it is not necessarily so; the proportion could be different. This proportion depends on the so-called “measure of existence”. For example, if an uploaded mind has a measure of existence of 0.01, and a real biological mind has measure of 1, I have a 100 to 1 chance to be a biological mind.

One of the most plausible candidates for measure of existence is the amount of energy used for calculations. The reason could be illustrated by Yudkowsky’s thought experiment with Ebborians—two-dimensional minds which could be split in the third dimension (E. Yudkowsky, 2008b). If real, such minds could be split until they reach a Planck limit of computation energy—basically, the Landauer limit (Bennett, 2003)—the smallest amount of energy needed to perform computations. Any computations that are more energy-consuming could be regarded as the sum of many Planck-level computations, or if we speak about minds, as a sum of copies. Thus, it is logical to measure the number of copies as the number of Plank computations.

Most BBs, even if they are functionally similar to the human brain, use very little or even “borrowed” (in the case of quantum fluctuation) energy for computation, as they are fluctuations, and smaller fluctuations are hyper-exponentially more probable. While the energy efficiency of human minds is high, it seems that a brain’s energy consumption is several orders of magnitude above the Planck level of computation.

This measure doesn’t work for mathematical minds. In the mathematical world, any object exists just once—for instance, the number 3 exists as just one number. However, small variations could be an analogue for measure here: that is, minds which have many small differences, like a few pixels difference, but the same main traits, like a name, could be regarded as many different copies of the same mind, thus having higher measure. In other words, minds with more low-level entropy will have higher measure in the mathematical world. This may be practically the same as Ebborian minds, as such minds also have small differences between layers.

## 4.3. Extremely simple or extremely complex BBs should dominate

We suggested above many ways in which BBs might appear. But if BBs are truly random, the way they appear doesn’t matter; they will be internally almost identical, and will not know anything of the outside world.

Another general idea is that a simpler BB should numerically dominate over more complex ones, and the distribution is very steep. If a coin toss is used to generate a mind, then a 1024-bit length mind will be more probable than a 2048-bit length mind, and the chance of creating the 2048-bit mind will be 1:21024 more likely than that of creating a smaller mind via thermodynamic fluctuation.

This seems to be a strong argument against the fact that I am a BB, as I am not the *simplest observer moment capable of writing the current text.* Iknow this because I can also see complex objects in my room (assuming that I am actually seeing them, not just claiming that I see them—but in this case, some longer thinking inside the OM is required).If I were blind, I still would be able to dictate this text, but I am not, so my observer moment is unexpectedly complex compared to the set of all observer moments who think about BBs.

This is not the case for mathematical BBs, which all have equal weight. However, if mathematical BBs are real, I should be surprised that I am too simple, as infinitely complex AIs should dominate (in the same way as there are more longer numbers than short numbers, e.g. only 10 single-digit numbers, but millions of 6-digit numbers). This argument was explored by Pereira in his Super-Strong-Self-Sampling Assumption (SSSSA), in which he argues that more complex minds have higher measure, but as I am not such a mind, then superintelligent minds do not exist (Pereira, 2017).

Pereira’s SSSSA is based on the idea that consciousness has “thickness”, and it is more probable that I am in the more complex moment of experience. This explains why I am not a cow or any other animal with a narrower consciousness. However, many animals may have approximately the same number of different qualia of colors and pain as humans, because vision is a very old adaptation, so if we measure consciousness as the sum of simultaneously experienced qualia, some animals may have the same amount of consciousness as humans, and given the very long history of animal kingdom, it will be again surprising to find oneself as a human, not a wolf.

In my “Meta-doomsday argument” article (Turchin, 2018), I suggested an alternative explanation: we are randomly selected from the class of beings who are asking the question about the Doomsday argument (the DA-aware reference class). Let’s illustrate it by a trivial example: I have an apple in my hands and ask my friend why I have an apple. He answers: if you have an orange, you would ask why you have an orange, so having an apple is a necessary condition for asking why you have an apple. In other words, one’s ability to ask the question is a necessary condition for the answer. In that case, my location in human history is not a surprise: only in the beginning of 21st century did the discussion about BBs become active. However, what about an upper limit? Do superintelligences still spend time thinking of the same questions, or it is just a “childhood activity”, which is done only at some specific level of personal development, in the same way as teenagers go into abandoned houses at night for the thrill of it—typically once in a lifetime.

The only ways to save the BB idea are either BB-typewriters or BB-chains of observers. Both create a more-or-less stable illusion of the real world, but they are not the original BBs. In some sense, they are a type of hostless simulation (and any hostless simulation is a type of BB).

## 4.4. A map of possible observers and BBs in the universe

In this section we classify all suggested ideas about freak observers based on two parameters: the size of the mind of an observer, and the reality of its existence. Real human on Earth are most complex and most real observers, and pure strings of numbers are most simple ever suggested observers.

|  |  |
| --- | --- |
|  | **Simpler minds More complex minds** |
|  |  | *Simple observer-moments* | *Structured observer moments* | *Functional model of the whole mind* | *Real brains in the real world* |
| **More mathematical****More physically real** | *Pure mathe-matical universe* | Strings of possible numbers | OMs as convolution netsOMs as other computational structures | Possible minds in mathematical universe | Modal realismPossible AI in mathematical universe running simulations |
| *Mathema-tical relation between real objects (dust theory)* | Qualia-based BB | Ordered, qualia-based BB | Boltzmann simulations;Chains of observer-moments | Minds in ancestor simulations |
| *Fluctuations in vacuum* | Quantum fluctuations producing a single OM;Panpsychism: any causal process has consciousness | Functional model of a mind appearing from a fluctuationDiffuse gas clouds BBs | Biological mind from amino acid soupBiological mind from a quantum fluctuation in Minkowsky vacuumQuantum random mind generatorRandom computations in real computersBlack-hole powered computer | Life on Earth as a fluctuation with some Darwinian selection; |
| *Real processes which use randomness to generate minds* | Thermodynamic observer moments (rock minds) | Longer BB, which consist on only two subsequent OM: 2OMs-BB |  | Solar system size BBsDarwinian selection generates real people on Earth |

## 4.5. Escaping the BB curse: BBs as reflections of real observers

The most unpleasant part of the BB idea is that if I am a BB, I should disappear in the next moment of time, that is, die. However, imagine the following: I am walking on a street and there are many mirrors on the buildings around me. There are hundreds of reflections of me in every moment, which appear and disappear with my every step. This doesn’t mean my death.

The alternative to the idea “I am a Boltzmann brain” is the idea that information doesn’t have “location”: I am located everywhere where my information is located. In that case, I am not randomly selected from among all of my copies, but I *am all* of my exact copies. The idea of “location” has meaning only to physical objects that could be located in different places: I am in Paris today, but was in Rome yesterday. But for the exact copies of observer moments, that doesn’t have any meaning: the “me” that was in Rome yesterday is different than the me in Paris. All copies have exactly the same information about their location.

Another example is the properties of numbers. 121 = 112, no matter if it is 121 matches, or 121 apples, or whatever else. The number 121 is 121 in all its copies, and its numerical properties don’t depend on the “location” of 121: in an apple box, or in a matchbox. In other words, 121 is located in the number space, but not in physical space; physical space doesn’t matter for 121, is it real, mathematical, will not appear or disappear. Yampolskiy explored (Yampolskiy, 2015) the space of possible mind designs, which is the space of possible Turing machines, and each Turing machine program could be presented as string of numbers.

The idea that “I am all my copies” helps to escape the BB curse, as while some of my copies disappear, others persist, and the disappearing copies doesn’t matter—even the total number of copies doesn’t change. This is similar to the idea of *big world immortality*: any BB is immortal, as there will be a real observer somewhere which will be its next moment of existence (Turchin, 2018). Moreover, for any freak observer, the existence of a real observer with the same experiences is possible.

## 4.6. Stable attractors for chains of observer moments

However, if only the real world exists, the natural next moment of 121 is 122. Now we have a question, how does the existence of BBs affects the distribution of the most likely next moments? If there are too many random BBs by measure, then the next moment will be chaotic, or at least strange.

A possible answer is that there will be something like Darwinian evolution of the chains of observer moments, which will eventually end in relatively stable chains, and such stable chains will be attractors of the unstable chains in the space of possible minds. For example, dreams look like a chaotic, unstable chains of events, but they eventually collapse into much more stable daytime experiences. Moreover, it is possible to suggest that dreams “were needed” in the Darwinian evolution of the chain of events (which has not happened in real physical time) in order to capture strange chains and attract them to the more stable chains.

## 4.7. Superposition of interpretations

The same observer moment could occur simultaneously in different BBs, in a real world and in a computer simulation—and maybe even in some other types of environments we can’t imagine now.

From this is observation follows the rather obvious *weak principle of superposition of interpretation*: any world could be real, a BB-illusion and a computer simulation. Let’s also note that the number of classical vacuum BBs is infinitely larger than number of simulations in the simulation argument, which itself is typically estimated as around 10-20 orders magnitude greater than the number of “real minds” (Bostrom, 2003).

However, we could also suggest a *strong principle of the superposition of interpretations:* the observed world actively prevents an observer from learning the observer’s location*.* In a simulation, such a principle has rather obvious reasons: the owners of the simulation don’t want the simulants to know for sure that they are in a simulation, and so they suppress any evidence of simulation, and even thoughts on the topic*.* This principle is very speculative, as actually the observed world implies my current OM is located in a real world.

In the case of BBs, the strong superposition of interpretation could best be explained via anthropic reasoning. If an OM knows for sure where it is located, it belongs to a much smaller class of OMs, making an OM less likely to arise in the first place. In other words, it may appear the superposition of interpretations prevents an OM from learning its exact location. (This is similar to Bostrom’s Adam and Eve thought experiment (Bostrom, 2001)). In such superposition of interpretations, there is a limit to which extent I could define my location based on observations, something similar to the Heisenberg uncertainty principle: the better I know in which world I am, the shorter my existence in it. This is a very tentative suggestion which needs additional considerations. The similar reasoning could be also applied to personal identity considerations: the more narrowly I define which pattern is me the sooner it will end; if I identify for infinitely many patterns I will exists as long as some of them continue to be.

## 4.8. Non-locality of the mind in the multiverse as a way to escape BBs

The idea that we could describe *a mind’ location inside a complex multiverse* may be itself contradictory.

Yampolskiy joked about it: “Occam’s Nightmare: an illusion of you is clearly just a meditative dream in a hypnotized Boltzmann brain-in-a-vat-of-drugs produced by a quantum fluctuation in a simulated universe created 5 minutes ago by an alien Superintelligence designed by a omnibenevolent God standing on a zombie turtle made by schizophrenic intelligent designer, and yes it is fractal turtles all the ways in 42 dimensions of holographic multiverse or so the hallucinating evil demon wants “you” to experience” and Michael Michalchik added “You forgot about the little homunculus inside the brain pineal gland during the true observation of the theater of the Mind creating the qualia who in turn has his own pineal gland with its own homunculus watching the homunculus is theater of the Mind, Ad infinitum”.

This joke underlines the internal absurdity of the idea of “mind’s location”. The idea of *physical location* is applicable only to physical objects, like brains, but not to digital entities. In other words, the following statements are true: “three chairs are located in the room”, and “number 3 is located between 2 and 4”. And when we start to think about BBs we start to think about minds not as about physical objects, but as of mathematical objects, and mathematical objects can’t have locations in the physical world: we can’t say that “number 3 is located in the room.”

However, one may say: while in *ontological sense* mind’s location is absurd, it still can have *practical sense* as the mind’s location could be used to predict next experiences. For example, if BB are dominating, then the next experience should be very chaotic (or non-existence, but it could be ignored). But we could discuss only “past next experiences”, so any predictions about real next experiences are unfalsifiable.?

To solve this problem Wei Dai suggested UDASSA: “Universal distribution + absolute self-sampling assumption”, summarized by Christiano (Christiano, 2011). The theory was never formally published as Wei Dai later turned to decision theories in order to solve self-location believes. “Universal distribution” here is based on computational complexity of describing the mind as well as its location in the multiverse, as “ASSA” means that I am randomly selected from all similar observer moments. UDASSA gives very low probability to BBs, as they are located in very remote places. The same way, to find a coherent book in a Babylon library of all random books one needs its address as long as the length of the book. UDASSA assumes that the physical multiverse *actually exists* AND that “address of a mind” matters. The second requirement is the weakest, as if there are two copies of me in two different boxes, I have equal chances to be any of them, no matter how far the boxes are from each other; in other words, the location of the boxes doesn’t affect my self-location believes. Mueller created a theory, similar to UDASSA, in which there is no postulated physical universe and no measure penalty based on the physical location. The probability distribution is based only on the complexity of function which describes transition between similar OM (Mueller, 2017). As only OMs are real in Mueller’s theory, there is no problem with BBs in it.

If we want to preserve the idea of actually existing multiverse, but remove the “location penalty” by using the “non-locality of mind” idea, we should remove “UD” from UDASSA and replace it with another probability distribution based on some properties of OM’s computational engine, like energy used for computations (more energy = higher measure). The reasons for it is discussed in our article about probability of being in a simulation (Turchin, Yampolskiy, Denkenberger, & Batin, 2019). In that case, most BBs use very small amount of energy for computations compared with real brains, but the sum of all BB’s measure could be still high. Anyway, only continuous BBs, which include several OMs, could contribute to the observables expected distribution, but such BBs are extremely less likely than OM-BBs and produce much less random expected observation.

In short, if I am located everywhere where my copies are present, then the disappearing of some of my BB-copies will not have observable consequences. It is not clear, if the extreme randomness of the next experience is also penalized in such setup. Imagine a BB, which consists of only two subsequent OMs: first is like my normal current OM, and the second is very chaotic, and after these two observer-moments, this BB disappears. Such 2OM-BBs will dominate between all long BBs.

## 4.9. Biological evolution as a shadow of my existence

One of the ideas behind BBs is that *everything possible exists*. I am possible, and thus I exist. And since I exist, there should be a most probable way to explain my existence. Because it has been observed, biological evolution is the most probable way my brain could appear from nothing.

This seems rather tautological, but it assumes that Darwinian selection is not the only mechanism of evolution. In some critical situations, evolution has “jumped” via random events; abiogenesis could be such event. If a first cell (or first self-replicating RNA) self-assembled completely randomly, everything else leading to a BB could happen via Darwinian evolution. Similarly, anthropic pressure could shape the directed nature and the speed of evolution to create observers (Turchin & Denkenberger, 2019). Scott Aaronson explored the question of evolution and complexity: is evolution a simpler way to generate complex observers than random generation? (Aaronson, 2012). The answer seems to be yes, but how could we know for sure? Yampolskiy et al wrote that in many cases evolutionary algorithms stack at some level of complexity (Mishra, Gupta, Chaturvedi, Shukla, & Yampolskiy, 2015). If this is true, then we exist as the result of a combination of processes of evolution and random mind generation.

# 5. Practical application of BBs

Both the existence and non-existence of BBs could be used practically. If BBs don’t exist, this could be used as an evidence for limiting the space of possible cosmological models and predicting the ultimate fate of the universe. If BBs do exist, this fact could be used to reach some form of immortality, or even as a counterfactual instrument to change the observed probabilities of events.

## 5.1. No BBs means a “Big Rip” soon, and astronomical waste for future supercivilizations

Our experience tells us that we are likely not classical BBs: our existence is ordered, not too simple or too complex, as we would expect of random minds—with all caveats about BBs’ inability to make true statements. In addition, the observed large size of the universe implies that we not a small fluctuation in space-time.

In other words, a lack of BBs rules out the heat death of the universe, but favors theories with a cut-off soon, like a “Big Rip” (Ćirković, Sandberg, & Bostrom, 2010). During Big Rip the universe is destroyed by its quick acceleration; however, some forms of quick acceleration could end up with de Sitter vacuum which is still capable to generate BBs. Linde et al find that the proportion of BBs created by de Sitter vacuum nucleation to real observers could be finite (De Simone et al., 2010).

In other words, no BBs implies a shorter period for our existence as a possible supercivilization, as the Big Rip will happen, by some estimations, 20 billion years from now. In a case of the heat death universe, we would have several orders of magnitude more time until the last red dwarf stars would burn out. In other words, the Big Rip is an astronomical waste of opportunity for future civilizations (Page, 2008).

The fact that I am not a random brain in amino-acid ocean is an evidence that biological evolution is not a *very* rare event, and empty planets have an upper level in frequency. This is also evidence against a very strong anthropic shadow, like false vacuum decay every 100 years, as in that case an observer could be randomly born from a non-sentient animal, like a dinosaur. More about expected frequency of events like false vacuum decay see in (Ćirković et al., 2010; Turchin & Denkenberger, 2019). Human DNA is around 50 per cent different from dinosaurs’ DNA, which is around 2 billion of bits of difference and while all of these mutations happening simultaneously is vanishingly unlikely, with a factor of 102E9, it is still more possible than many other BBs.

The BB argument should be applied not only to the observable universe fate, but to the whole multiverse. If there are two similar universes in the multiverse, one with early cutoff, and another with Heat death and many BBs, Self-Indication Assumption principle, SIA, (Bostrom & Cirković, 2003) implies that we should be in the second one with the heat death.

## 5.2. Boltzmannian immortality

Loew argues “that if survival of death is possible at all, then we almost surely will survive our deaths because there almost surely will be Boltzmann duplicates of us in the distant future that stand in appropriate relations to us to guarantee our survival” (Loew, 2017). Loew also argue that there will be many BBs in the future of infinite universe which are not only copies of my mind, but causally connected with it “in wide causality view” (Parfit, 1984). This is similar to the causality of the butterfly effect: any past event affects the entire future light cone, including all random fluctuation in it. Thus, if in the future there appears a BB which represents a moment of my life, which is a continuation of my mind state after death in this world, it will be also causally connected with my mind, thus ensuring continuity of consciousness, if such causal continuity is needed for personal identity.

Boltzmannian immortality is a particular case of many-worlds immortality I described in the article “Forever and again” (Turchin, 2018). If probabilistic resources are infinite then all moments of you exist all at the same time. So it is not a sequence of moments, but all possible moments representing you at all times. However, there is an illusionary time based on internal clocks of each moment, in other word, the moments could be numbered and could have an illusion of sequence of moments. In Boltzmann Immortality (B-immortality), after the first BB which presents the first moment of my experience after death (at moment t+1), there will be a second BB, eons later, which will represent the continuation of the first BB, (at moment t+2), etc. Thus, B-immortality implies the possibility of stable chains of observer moments, and that such chains will be not drastically different than the normal experience of a mind, or, otherwise, such immortality will collapse into a kaleidoscope of strange and unpleasant experiences.

## 5.3. “Permutation City” immortality

B-immortality could provide some form of immortality, but there is no guarantee that it will be good in any meaningful sense. In the best case, it will look like just miraculously escaping death, but experiences of injuries and semi-consciousness would dominate.

However, B-immortality could be augmented in a way which prevents bad outcomes. One way to do this is described in Egan’s novel *Permutation City* (Egan, 2010). The main idea here is that B-immortality is applicable not only to biological minds, but also to computers.

The next state of a computer is completely deterministic relative to its initial state, so the B-immortality of a computer will not create strange kaleidoscopic computer states, but will create a stable chain of the computer’s states. If an uploaded mind is running on such a computer, such an upload will also escape experiencing of the bad kaleidoscopic experience as BB. So, the whole algorithm of computer-assisted B-immortality in Permutation City works as follows:

1. An operating system is created that has the capability to grow infinitely given sufficient computational resources. In the novel, such a system is based on something like cellular automata. But inside this operating system any other computer programs could exist as a virtual machine, including a mind upload. (In *Permutation City* the whole setup uses an initial state which can’t appear via natural evolution of cellular automata, but it is not clear if this is a necessary condition).
2. A human mind is scanned and uploaded into the system.
3. The whole system, after testing and running for a short time, instantly turns off.
4. The next state of the whole computer could be only a Boltzmann brain (or some other form of many-worlds immortality).
5. There is an infinite stable chain of computations following from this initial BB, and each of them is another BB. The BBs in this chain are immune to any outside dangerous physical processes and thus could be immortal. (There could be some internal computational processes that may still look like aging and death, e.g. bug accumulation and/or program halting, and this problem would need to be solved for true immortality.)

## 5.4. Exploiting glitches in the chains of observer moments

The main principle here is that it is not a bug, but a feature. If there are some glitches in observer chains, they could be exploited to create some form of “magic”, instruments to manipulate probabilities which are above main physical laws, and that may look like manipulating the number of coincidences. Quick changes in one’s perception, maybe under effect of drugs, may increase the probability of such irregularities.

## 5.5. Flux universe

The idea of a flux universe is that everything outside my current experience is in a state of uncertainty. The same observer moment could belong to many different personalities. If I don’t currently keep the name of my mother in my working memory, her name is uncertain relative to the moment now, as the same OM could be in personalities with mothers with different names.

Along the same lines, a person on the LessWrong blog (Eitan, 2015) has expressed concerns that if he goes to sleep, he will forever “lose” his native personality, as his OM chain will drift away in the half-dream state, and he eventually will be someone else. (In Nabokov’s short story *Ultima Thule* a character learns some cosmological secret, perhaps, that he is just a BB, and start act erratically, destroying his own life. This could be good fictional illustration of BB or flux universe idea if it will be taken seriously.)

The idea of such a flux universe is closely connected with BBs, as if chains of BBs are real, only the content of consciousness has any predictive meaning for the next experiences—not long-term memory or any unknown and hidden facts.

This could be used for counterfactual probability manipulation. If I know very rare fact X about me (like my rare medical diagnosis), and I remove this knowledge from my current consciousness, then I become a member of the class of observers, most of whom don’t know fact X, and thus I escape situation X (i.e. most likely will not have this diagnosis). While this is absurd in the real world, it is completely logical for a world in which observations are real, but not the external world. Obviously, it can’t be experimentally tested, as I must even forget that I was trying to escape the knowledge of fact X.

## 5.6. BB uncertainty and the problem of constructing of a global prior

In Bayesian reasoning the global prior is an initial world model; but if freak observers dominate, everything is possible, and strange events are not evidence for anything. In that case, science becomes impossible, but it is possible in our world. Bostrom wrote about this problem in case of modal realism (Bostrom, 2000) ????.

# Conclusion

While the existence of “normal” BBs is either unlikely or irrelevant, BBs with some ordering may have observable consequences. There are two types of such BBs: Boltzmannian typewriters (including Boltzmannian simulations) and the chains of observer moments. BBs may also have practical applications: for measuring the size of the universe, achieving immortality or even manipulating the observed probability of events.

# Literature

Aaronson, S. (2012). Why philosophers should care about computational complexity. *In Computability: Gödel, Turing, Church, and beyond (Eds*. Citeseer.

Albrecht, A., & Sorbo, L. (2004). Can the universe afford inflation? *Physical Review D*, *70*(6), 063528.

Almond, P. (2003). *Indirect Mind Uploading: Using AI to Avoid Staying Dead*. Retrieved from https://web.archive.org/web/20120310211305/http://www.paul-almond.com/IndirectMindUploading.htm

Argonov, V. Y. (2012). Neural Correlate of Consciousness in a Single Electron: Radical Answer to “Quantum Theories of Consciousness.” *NeuroQuantology*, *10*(2). Retrieved from http://neuroquantology.com/index.php/journal/article/view/548

Armstrong, S. (2018). Are you in a Boltzmann simulation? - LessWrong 2.0. Retrieved February 4, 2019, from LessWrong website: https://www.lesswrong.com/posts/ygELzNSAF5nzLXD7j/are-you-in-a-boltzmann-simulation

Bennett, C. H. (2003). Notes on Landauer’s principle, reversible computation, and Maxwell’s Demon. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, *34*(3), 501–510.

Bostrom, N. (2000). Observer-relative chances in anthropic reasoning? *Erkenntnis*, *52*(1), 93–108.

Bostrom, N. (2001). The Doomsday Argument Adam & Eve, UN++, and Quantum Joe. *Synthese*, *127*(3), 359–387.

Bostrom, N. (2003). Are You Living In a Computer Simulation? *Published in Philosophical Quarterly (2003) Vol. 53, No. 211, Pp. 243-255.*

Bostrom, N., & Cirković, M. M. (2003). The Doomsday Argument and the Self–Indication Assumption: Reply to Olum. *The Philosophical Quarterly*, *53*(210), 83–91. https://doi.org/10.1111/1467-9213.00298

Carroll, S. M. (2017). Why Boltzmann Brains Are Bad. *ArXiv:1702.00850 [Astro-Ph, Physics:Gr-Qc, Physics:Hep-Th, Physics:Physics]*. Retrieved from http://arxiv.org/abs/1702.00850

Chalmers, D. J. (1996). Does a rock implement every finite-state automaton? *Synthese*, *108*(3), 309–333.

Christiano, P. (2011). The Absolute Self-Selection Assumption - LessWrong 2.0. Retrieved April 15, 2019, from LessWrong website: https://www.lesswrong.com/posts/QmWNbCRMgRBcMK6RK/the-absolute-self-selection-assumption

Ćirković, M. M., Sandberg, A., & Bostrom, N. (2010). Anthropic shadow: observation selection effects and human extinction risks. *Risk Analysis: An International Journal*, *30*(10), 1495–1506.

Crawford, L. (2013). Freak Observers and the Simulation Argument. *Ratio*, *26*(3), 250–264. https://doi.org/10.1111/rati.12009

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

Dyson, L., Kleban, M., & Susskind, L. (2002). Disturbing Implications of a Cosmological Constant. *Journal of High Energy Physics*, *2002*(10), 011–011. https://doi.org/10.1088/1126-6708/2002/10/011

Egan, G. (2009). Dust Theory FAQ. Retrieved from http://www.gregegan.net/PERMUTATION/FAQ/FAQ.html

Egan, G. (2010). *Permutation city*. Hachette UK.

Eitan, Z. (2015). The Consequences of Dust Theory. - LessWrong 2.0. Retrieved May 17, 2019, from https://www.lesswrong.com/posts/is7ieoWyiyYRc7eXL/the-consequences-of-dust-theory

Furstenberg, H. (1981). Poincaré recurrence and number theory. *Bulletin of the American Mathematical Society*, *5*(3), 211–234.

Loew, C. (2017). Boltzmannian Immortality. *Erkenntnis*, *82*(4), 761–776.

Mishra, M., Gupta, V., Chaturvedi, U., Shukla, K. K., & Yampolskiy, R. V. (2015). A study on the limitations of evolutionary computation and other bio-inspired approaches for integer factorization. *Procedia Computer Science*, *62*, 603–610.

Mueller, M. P. (2017). Law without law: from observer states to physics via algorithmic information theory. *ArXiv:1712.01826 [Physics, Physics:Quant-Ph]*. Retrieved from http://arxiv.org/abs/1712.01826

Page, D. N. (2008). Is our universe decaying at an astronomical rate? *Physics Letters B*, *669*(3), 197–200. https://doi.org/10.1016/j.physletb.2008.08.039

Page, D. N. (2017). Bayes Keeps Boltzmann Brains at Bay. *ArXiv Preprint ArXiv:1708.00449*.

Parfit, D. (1984). *Reasons and persons*. OUP Oxford.

Penrose, R., & Gardner, M. (2002). *The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics* (1 edition). Oxford: Oxford University Press.

Pereira, T. (2017). An Anthropic Argument against the Future Existence of Superintelligent Artificial Intelligence. *ArXiv:1705.03078 [Cs]*. Retrieved from http://arxiv.org/abs/1705.03078

Tegmark, M. (2014). *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality* (1st edition). New York: Knopf.

Tomasik, B. (2019). What Are Suffering Subroutines? Retrieved May 26, 2019, from Reducing Risks of Future Suffering website: https://reducing-suffering.org/what-are-suffering-subroutines/?fbclid=IwAR2y8eP71cCNF6YL5uICmEIV04n\_W8XzjdkcCuAq8cKZ9PYlsMEZK2eXoBc

Turchin, A. (2018). *Forever and Again: Necessary Conditions for the “Quantum Immortality” and its Practical Implications*.

Turchin, A., & Denkenberger, D. (2019). *A Pin and a Balloon: Anthropic Principle Implies Underestimation of the Fragility of Our Environment and a Higher Risk of Runaway Global Warming*.

Turchin, A., Yampolskiy, R., Denkenberger, D., & Batin, M. (2019). *Simulation Typology and Termination Risks*.

Turchin, Alexey. (2018). *A Meta-Doomsday Argument: Uncertainty About the Validity of the Probabilistic Prediction of the End of the World*. Retrieved from https://philpapers.org/rec/TURAMA-4

Turchin, Alexey, & Yampolskiy, R. (2019). *Glitch in the Matrix: Urban Legend or Evidence of the Simulation?* Retrieved from https://philpapers.org/rec/TURGIT

Yampolskiy, R. V. (2015). The space of possible mind designs. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9205*, 218–227. https://doi.org/10.1007/978-3-319-21365-1\_23

Yudkowsky, E. (2008a). How An Algorithm Feels From Inside - LessWrong 2.0. Retrieved May 17, 2019, from LessWrong website: https://www.lesswrong.com/posts/yA4gF5KrboK2m2Xu7/how-an-algorithm-feels-from-inside

Yudkowsky, E. (2008b). Where Physics Meets Experience - Less Wrong. Retrieved February 8, 2018, from LessWrong website: http://lesswrong.com/lw/ps/where\_physics\_meets\_experience/

Yudkowsky, Eliezer, & Soares, N. (2017). Functional Decision Theory: A New Theory of Instrumental Rationality. *ArXiv:1710.05060 [Cs]*. Retrieved from http://arxiv.org/abs/1710.05060