Existence, consciousness, and ethics: Extending the Mathematical Universe Hypothesis

Mads J. Damgaard*

January 23, 2024

Abstract

We give some arguments for why the Mathematical Universe Hypothesis (MUH) might be too restrictive in its assertions of what can exist, and that the universe/multiverse might be formed by more than what can be expressed mathematically. In particular, we show a thought experiment which indicates that the principle of materialism in general is an inadequate hypothesis of how consciousness appears. Instead we propose a novel approach to solving the problem of consciousness, which is to hypothesize that each universe might have different laws of consciousness, just as they might have different laws of physics. We also show that MUH, without such consciousness laws, and without any other modification, leads to complete chaos for the observers in the multiverse, and therefore cannot hold as it is. We then go on to propose another theory of existence, which does not seem to lead to complete chaos. This theory hypothesizes that the multiverse is the consequence of what we might call the Logic of Everything (LoE) continuously deducing new truths about itself. Lastly, we briefly discuss what fundamental ethics can be derived from this theory, and others like it.

1 Introduction

A complete theory of existence would need to answer two fundamental questions in particular: What is the set of things that *can* exist, and out of that set, what is the subset of things that actually *does* exist?

The Mathematical Universe Hypothesis (MUH), put forward by Tegmark [1–3], is a remarkable example of what we might refer to as a *Theory of Anything* (ToA), in the sense that it states that 'everything that *can* exist *does* exist.'[1] MUH then further states that everything that *can* exist is *mathematics*, and thus that the multiverse is equivalent to the collection of all mathematical structures, and all mathematical truths.

In this paper, we will put forward some counterarguments to this theory, by arguing that mathematics might be insufficient for describing abstract concepts such as consciousness fully. This will lead us to propose an extension of MUH as a theory where one allows for more than simply mathematics to exist. We will then discuss some dangers for theories of existence, such

*B.Sc. at the Niels Bohr Institute, University of Copenhagen. B.Sc. at the Department of Computer Science, University of Copenhagen. E-mail: fxn318@alumni.ku.dk.

[1] Steinhardt [8] uses this term, 'Theory of Anything,' about theories of multiverses in general, albeit in a somewhat deprecatory way. However, there is no reason why we could not repurpose the term and use it, in a more neutral way, to denote theories which hypothesize this profound symmetry of existence, i.e. that 'everything that *can* exist *does* exist.'

as MUH, of running into a problem of what we will refer to as *complete chaos*, which we will take to mean that the experience of an average observer in the multiverse is completely chaotic, and that no real predictions can be made about the future of any observer, given their past. And finally, we will propose a hypothesis of what we will refer to as a *self-deducing Logic of Everything* (LoE) as a reasonable candidate for a theory of existence, which does not seem to necessarily lead to complete chaos.

The theories proposed in this paper are generally not meant to be thought of as theories of physics, as they do not lead to any significant predictions within our own universe that we can go out and test, at least not at this stage, anyway. Rather they are meant as theories of metaphysics. Their value mostly lies in how they might pique our curiosity, and potentially satisfy some of it, as well as, not least, how they might make us think about our own place in existence, our consciousness, the question of 'what comes after death,' and about what actions are ethically right and wrong.

2 The potential limitation of even natural languages to fully define a theory of existence

In this section, we will try to argue that mathematics might be inadequate to encapsulate what it means to *experience something* (consciously), and what makes different sensations feel the way they do.

This is a rather abstract point, but it is made less abstract if we look at particular sensations that are somewhat related and ask ourselves what would happen if one sensation was exchanged with another. For instance, if we take the experience of feeling something hot versus the experience of feeling something cold, we know that these experiences create different internal sensations in our mind, make us think different thoughts about them, and make us react differently to them. However, if one were to exchange the two feelings from birth, would we ever know the difference?

The same thing can be said about the experience of seeing the color red versus the color blue. If these colors were switched[2] from birth, would we ever know the difference? Perhaps not. Maybe we would still have exactly the same feelings towards the color red, even if we saw it internally as blue instead. Maybe we would thus have exactly the same mental associations with it, and react to it in exactly the same way. And if that is the case, then the laws governing how we experience different colors consciously does not need to be part of the *physical* laws of our universe, as they would be inconsequential to the physical dynamics of the objects within it.

One might then ask: Would it even change the world if those colors were exchanged, then, or is the difference between seeing the color red versus the color blue purely an illusion on the subjective level? Well, some might argue that, indeed, it makes no existential difference to 'switch the colors,' and that such a thing is meaningless to begin with. Others, however, might argue that switching the experiences of how we see different colors in our mind *does* make a difference, even if all our other thoughts and feelings towards them (and our reactions to them) are unchanged.

The latter assumption has a rather intriguing implication. It means that none of us will ever be able to truly explain the experience of how we see a certain color in our mind. For say that you tried to formulate an explanation of how you see e.g. the color red internally. Then

[2] Of course, there are more colors in reality than just red and blue. However, the full set of colors can still be contained in a two-dimensional plane, and one could imagine making rotations and reflections in that plane.

in a hypothetical parallel universe, there might be a person exactly like you in every way who would formulate the exact same explanation, but who nonetheless sees the same color internally in the same way as you see the color blue.

This would mean that even natural languages are not strong enough to semantically define the internal experiences of seeing colors. And since the set of what can be defined semantically in mathematics is only a subset (proper or improper, depending on how you define 'mathematics') of what can be defined in natural languages, it would certainly mean that mathematics is not strong enough semantically to define the multiverse fully.

Therefore, if we want to keep everyone onboard, including people (e.g. perhaps the author of this paper) who thinks that there *is* a real existential difference in experiencing different colors, and that this difference cannot be formulated in any natural language, we should thus extent MUH as a theory, namely by leaving this matter unspecified. Anyone who wants the claim that color experiences *can* in fact be described in natural languages, or that the difference is really just an illusion on the subjective level, can then simply add that statement as a subsequent axiom of the theory.

3 Arguments against materialism

An important part of MUH is the *materialistic* assumption that consciousness appears automatically out of objects that are complicated enough to have a kind of *self-awareness*. On the surface, this seems like a reasonable assumption. However, in this section, we will look at a thought experiment which indicates that materialism might be an inadequate principle when it comes to theories of consciousness.

For this thought experiment, let us imagine that we have two identical neural networks next to each other. We then turn on one of the networks and let it run for a while, feeding it sensory input while it runs. Then we do the same for the other network, giving it the same input. Imagine further that the networks use analogue computations that are not completely deterministic, and that there is a very high likelihood that the computations/thoughts of the two neural networks will diverge after a short while, even when they are given exactly the same input. Surely this would result in two distinct conscious experiences, would it not?

For the rest of the thought experiment, we then go through a process of gradually gathering these two neural networks into one. First we gather them in time, making their runtime overlap more and more until they are run simultaneously. Then we gather them closer and closer in space, until each neuron and each neural link is exactly adjacent to each other, almost touching. Then we start coupling the neurons slightly such that, although the computations still might diverge, each neuron now has a higher chance of firing when its neighbor does. We could also even couple the memory of the networks slightly such that at any point in time, there is a small change for a memory cell to erase its own value and copy its neighbor's value instead. This would thus give a higher chance of the two computations to converge again after having diverged from each other. When this coupling is very weak at first, the networks will thus still practically function independently of each other. However, when the coupling is turned up high enough, the two neural networks will practically become one. And once this is the case, surely this combined network would result in only one conscious experience, would it not?

Now, most people would probably agree with both these statements, i.e. that there are two conscious experiences at first and then only one at the end when the two neural networks have been gathered into one, at least if they also accept the initial premise that these networks *can* result in conscious experiences in the first place. But the big question is then: Where in this process does this change? When exactly do we go from having two conscious experiences to having only one? It must happen somewhere in the middle, but *when*?

A quick solution to this question would be to state that there must exist a threshold of how strongly two neural networks can couple to each other before they no longer hold two separate consciousnesses. But once you have chosen a specific threshold, you could always ask: Why exactly that one? Why not the threshold that is a microscopic degree higher? This arbitrariness makes it hard to believe that there is a fundamental logic in the entire multiverse that makes consciousness appear automatically out of objects that are complicated enough.

A better solution seems to simply leave the theory open towards having different *laws of consciousness* for each universe, by saying that not all universes necessarily have to have the same laws for when consciousness appears. In some universes, the threshold that we have just discussed might be one thing, in others it might be something else.

But hold on, could the difference between when we have one or two consciousnesses not simply be dependent on whether or not the two neural networks make the exact same computations? If they make the same computations, then they result in one conscious experience, and if they diverge, then they result in two? Well, the problem with an assumption like this is that it might very well lead to 'complete chaos.' For if all brains who make the exact same computations lead to the same conscious experience, counting as only *one* observer, it means that all conscious experiences will happen only once in the universe (or in the multiverse, depending on your assumption). If then the universe (or multiverse) is only big enough, it means that all imaginable conscious experiences happen once and only once. But this results in an even probability distribution for all possible experiences. So if you for instance imagine that you drop a ball on a table, the experience of seeing the ball tunnel through the table and appear on the other side is as likely as seeing it bounce back up. And the experience of seeing it move sideways when you drop it is as likely as seeing it move down. Thus we effectively get a universe (or multiverse) of complete chaos in the eyes of the observers.

Now, we cannot categorically rule out that there might exist some profound principle of when consciousness appears, with no free parameters and no arbitrariness to it, which one might then propose as a universal theory of consciousness. But the thought experiment that we just discussed does make it seem rather unlikely. And until such a beautiful, parameter-free principle has been found, we might as well extend our theory of existence to allow for the fact that not all universes necessarily have to have the same laws of consciousness. Then if someone at some point *does* find a beautiful principle of consciousness that they wish to propose as a universal (or "*multi*versal") principle of all existence, then they can just add that as a subsequent axiom of the theory.

4 An extension of MUH

To summarize the results of the last two sections, we have thus argued that MUH might need to be extended as a theory in order to encapsulate all aspects of the multiverse, and that there might exist more than just pure mathematics. Specifically, we have argued that the concept of experiencing various sensations might not be fully explicable by mathematics, or even by what we can ever formulate in a natural language. And we have also argued that consciousness might not arise automatically out of self-aware objects, e.g. brains or machines, as the principle of materialism proposes. For it seems that a theory of consciousness would always need to determine some free parameters, and if the theory is not parameter-free, how can it be a universal principle throughout all of existence?

Well, if we stick to the ToA principle and say that 'everything that *can* exist *does* exist,' we should then simply hypothesize that there exists a version of the same physical universe for each possible choice of parameters of the given theory of consciousness, both in terms of when consciousness arises out of physical objects, and also for how observers experience different sensations internally. These choices will then result in what we could call different 'consciousness laws' for each universe.[3] These might then exist side by side with the physical laws, so to speak, functioning together to define all objects and all conscious experiences within the universe.[4]

Now, for some, this might seem like an unsatisfactory answer to these questions of consciousness, especially for anyone who has their hopes up that we might someday find a profound answer which explains all of what it means to be conscious. But as we have argued above, such an answer might not be obtainable in our universe. First of all, there might be some free parameters to the laws of consciousness that we have absolutely no way of determining empirically as observers within the universe. And furthermore, the full answer might not even be expressible in any logic/language that is available to us. For it is not a given fact that beings within a universe can necessarily understand everything outside of it, or about it.

With this proposal of an extended MUH as a theory of existence, we can then go on to ask the important question: *Why* does anything exist? Well, a very simple, yet reasonable approach to answering this question, which might satisfy some, is simply to say: The only thing a thing can do is either to exist or not exist, is it not? Now, we can already rule out empirically that things do not exist. And if we discount the option of all things *both* existing and not existing at the same time, since this is effectively the same as everything existing, we are thus left with the only option that all things exist.

That is, at least if all things are treated equally by the multiverse. However, as we will discuss in the next section, having all things being treated exactly equally by the multiverse, in terms of their existence and non-existence, actually leads to some problems.

5 Problems of complete chaos: An argument against the original MUH

We have already discussed how the assumption that all equivalent neural network activities lead to the same, *singular* conscious experience might lead to complete chaos, at least if the universe/multiverse is only big enough. In this section, we will discuss another potential danger that might make some theories of existence lead to complete chaos as well. And it seems that the original MUH, without the extensions proposed above, is one of those theories.

The argument is to first of all consider the fact that in quantum mechanics, one is always free to change the basis of the given Hilbert space via a unitary transformation. Thus, if ψ represents some quantum state, we are free to change it by

$$\psi \to U\psi, \tag{1}$$

where U is a unitary operator, as long as we also remember to change any operator, A, of any

[3] An interesting consequence of the possibility that different universes might have different consciousness laws is that a machine that is *not* conscious in one universe might be so in another. But since the physical laws for all we know will be independent of the consciousness laws, it means that we still always have to treat a complex enough machine that seems to be conscious as such. First of all, there will be no way for us to ever tell which universe we belong to, and second, even if we guessed correctly, the same actions that we take would then just be mirrored by other versions of ourselves in other universes, with the exact same physical laws as ours, but with different consciousness laws.

[4] One might even go a step further and hypothesize that the collection of conscious experiences of a universe *is* the universe, and that the objects of the universe does not exist outside of our consciousness, so to speak. The physical laws and objects could then simply be seen as a by-product of the universe defining some structure to its collection of experiences.

given observable by

$$A \to UAU^{-1} \tag{2}$$

as well. (This is similar to how one is free to make coordinate transformations in classical physics.) However, when there are no consciousness laws for the universe, and consciousness simply appears automatically from objects according to the principle of materialism, there is nothing that gives e.g. the position operator, X, priority over any given transformed position operator, UXU^{-1}, when it comes to determining the locations and the motions of the neural networks (and other self-aware machines) in the universe. This means that the original MUH, as described by Tegmark [1–3], will treat each members of the set

$$\mathbb{X} = \{UXU^{-1} \,|\, U \text{ is unitary}\} \tag{3}$$

equally.[5]

Now, the problem with this is that if we look at any given large volume of our universe and ask the question of what conscious beings live in that volume, then regardless of what we as observers observe in that volume, we need to sum over the ensemble of all possible quantum states that can be obtained from a unitary transformation when counting those beings. And the resulting ensemble will simply be the ensemble of maximal entropy. The probability of any observers within the volume of seeing a low-entropy universe like the one that we observe here from Earth, i.e. with stars all around that are alive and burning hydrogen, is thus extremely unlikely.[6] In fact, it would be much more likely for any observer that sees such a universe to actually be a brain in a vat floating in space somewhere, which has assembled itself from chaos out of pure coincidence, and where the brain is simply coupled to a simulation that makes it see such a universe.[7] The explanation for this is that such states, although they seem very unlikely, will still have much higher entropy than states with whole stars full of high-energy hydrogen.[8]

'Oh well,' one might say, 'then each one of us is perhaps just a brain floating in a vat somewhere in space, so what? If the simulation to which our brains are coupled is constructed randomly, then it might as well simulate the laws that we see in our universe as well anything else.' That is true, but since such a vat would be constructed out of pure coincidence, its lifespan would probably be quite short on average before disintegrating or malfunctioning. And therefore the universe has no business looking as old as it does. For instance, there would be no reason for us to find dinosaur fossils in the ground that can be carbon-dated back millions of years. In other words, there is no reason for us to observe a universe with a seemly consistent history of the past if that history never actually existed. This 'brain in a vat' hypothesis is thus quite unsatisfactory.

[5] One might point out that the physical laws will not generally remain the same after such a transformation, unless the U's commute with the Hamiltonian. However, we could also go a step further and say that, a priori, there is nothing stopping us from making time-dependent transformations of X as well, in which case we can not only transform ψ to any state that we want, but we can also get any time evolution that we want.

[6] This argument assumes that we can take some parametrization of the set of all unitary operators, or at least a subset thereof, and give this parameter space an even probability distribution. For instance, we might take the set $\{U(t) \,|\, 0 \leq t \leq T\}$, where $U(t)$ is the time evolution operator after assuming periodic boundary conditions for the given volume, and where T is a large enough time such that $U(T)$ approximates the identity operator (meaning that T is the time it takes the volume to essentially loop). Admittedly, we have not assumed any axioms which tells us that we *can* do all this. However, the whole point is that without any consciousness laws, or other modifications to the theory, there is nothing that tells us that we *cannot*.

[7] Page [6] discusses this problem, and how it relates to cosmology. In that discussion, the brains are not necessarily coupled to a simulation as described here, as they might also simply assemble themselves alone, outside of any vat, and just experience random sensory input.

[8] Or put in other words, the number of states with 'brains in a vat' will be much greater than the number of states with burning stars all around in the full ensemble of maximal entropy, which means that the likelihood of seeing the former, if you pick out a state at random, is much greater than that of seeing the latter.

This argument seems to disprove the original MUH without consciousness laws, or at least without any other specification of how to determine the locations of the observers.[9] However, the original MUH is not the only theory that faces the danger of leading to complete chaos. If we keep the ToA principle as part of the extended MUH, i.e. the principle that 'everything that *can* exist *does* exist,' then it too might lead to the same paradox. For even though the consciousness laws of a *particular* universe might then be defined specifically according to the operator X, there is nothing preventing other universes from having consciousness laws which are defined according to any other member of $\mathbb{X}$. And if all these members are treated equally by the multiverse, we thus still end up with the same paradox.

6 An ordering of the universes

The paradox described in the previous section arises from not having any order to all the (infinitely many) possible observers in the multiverse. If there is no particular order to these, and if we are thus free to order them however we want when calculating their prior probabilities, it is no surprise that we can thereby reach contradictions and paradoxes.

Thus, the only way to avoid this problem, it seems, for any ToA such as the extended MUH, is to include an axiom which gives all the possible observers in the multiverse an order, either a partial or a total one. When they have that, one can then assign a probability to each distinct conscious experience of any observer in a consistent way. And depending on the order, we might avoid the paradox of complete chaos.

The first thing that might come to mind is to then consider how the *universes* might be ordered among themselves. If we can find such an order, and preferable one that has some reasoning behind it, it would give us at least a partial order for all the observers, as they could then each be ordered according to the given universe to which they belong.

Now, if we assume that universes can include an infinite number of observers, we still run into problems when defining the probability calculation. However, let us simply put this issue aside for the moment and assume that there are only a finite number of observers for each universe. Then for the probability calculation, we can line up all the universes, first of all, and for each particular kind of conscious experience, we can find its prior probability by counting all observers with that experience in each universe up until a very large number, N. By then dividing with the total number of observers, we get the frequency of that particular experience in the multiverse. And if this frequency converges when N tends towards infinity, we thus hereby get the prior probability of that experience.

The big question is then: How could such an ordering of the universes come about? Well, one possible answer is the following. Consider the laws of the multiverse. Since the multiverse is 'all of existence' by the definition that we are using in this paper, these laws will thus be the fundamental logic of everything, of existence in its entirety. Let us refer to this as the *Logic of Everything* (LoE) for short. Now, we have made steps in this paper to argue that this LoE might be more abstract and complicated than simple mathematics, and perhaps even more complicated than anything that we can ever comprehend. However, it might still have some sort of structure to it. In particular, it might have an order in the sense that some things take less information to describe than others. Maybe we could even think of it as kind of "language," only where the fundamental "words" of this language might be far beyond our comprehension. And with this picture, all universes might essentially be ordered in terms of the sequence of "words" that define them. This could be a total order if the "words" themselves have an order

[9] In fact, it seems to disprove any theory which relies only on the principle of materialism to give rise to consciousness, and which at the same time gives no particular priority to some Hilbert space bases over others.

to them, or it could at least be a partial order in the sense that the "sentences" might be ordered in terms of the number of "words" that they contain.

This is exactly the kind of thing that could give the multiverse the order that we seek. If all universes are ordered according to the information it takes to describe them (in the LoE), then there might be a reason why universes that use the position operator X in their consciousness laws, without any transformation, would be more frequent than ones who use UXU^{-1} for some U, as it simply takes a lot more information when U needs to be defined as well. If we therefore sum over all universes up to some large number of "words," N, to compute the frequency of certain kinds of experiences, we see that for any given N, the number of "sentences" that use simply X for their consciousness laws will be much greater than the number of "sentences" that use UXU^{-1}. Thus, a theory of existence like this *could* give an explanation for the principle of Occam's razor, and why this is a valid principle for the observers (on average) within the multiverse.[10]

Now, the first potential objection to this theory is that we have not yet given any reasonable explanation for why there should be only a finite number of observers for each universe. And this would not only mean that space needs to be finite for each universe, but also time as well. These are rather strict assumptions. But it gets even worse than that, because, as we will see in the next section, even when space and time are assumed to be finite, we might still run into the problem of the theory leading to complete chaos, regardless.

7 A potential problem with unbounded time or space

If we assume that all the universes in the multiverse are ordered by the information it takes to describe them in the LoE, and then "played out" one at a time (or at least a finite number at a time), then with an additional assumption that each universe is finite, and thus includes a finite number of observers, we do get a way to calculate the prior probability for any given experience where the probabilities might converge.

However, if the time and space are unbounded, their extend then ought to be defined in the "sentences" as well. To then understand why this might be a problem, let us ask ourselves the following question: When we sum over all universes up to some large number of defining "words," N, what will be the average space and time of those universes (at least if we only look at universes where 'space' and 'time' are meaningful concepts)? Well, if we look at a universe like ours, the extent of space and time might be decided by free parameters. Imagine therefore a template of a "sentence" defining a universe of the general form

$$\text{'universe } A(p_1, p_2, \ldots, p_n) \text{ exists,'} \tag{4}$$

where A is a function returning a description of a universe, and $p_1, p_2, \ldots, p_n$ are the free parameters of that universe. If we then sum over all variants of this template where the total number of "words" is less than or equal to N, it means that all the free parameters are defined

[10] One might note that this conclusion assumes that when an object is repeated across different "sentences," it then exists several times in the multiverse, meaning that you have to count the observers within that object for each repetition. Or to put it in more concrete terms, let us suppose that the LoE includes conjunctions, i.e. includes something like the word 'and.' Then a sentence describing a universe can state something like: 'Universe A exists *and* universe B exists.' We then might assume that this will lead to what we could call a *local multiverse*, i.e. of two independent universes existing side by side as a sort of *sibling universes*. With the ToA principle, the multiverse will then also include the sentence stating that 'universe A exists,' as well as a sentence stating that 'universe A exists, and universe C exists,' etc. And if we assume that these three sentences, all including the clause that 'universe A exists,' leads to three different copies of universe A in the multiverse, it seems that we will then indeed get this principle of Occam's razor, as desired.

by up to $N - a$ "words," where a is the number of constant words in the template. Now, what is the average value of a parameter that is defined by $N - a$ or fewer words? Anyone who knows anything about very large numbers, such as e.g. Graham's number[11], will also know that even if the "language" of the LoE is quite simple, the maximal value will grow so incredibly fast that it will quickly become the dominant one in the calculation of the average. This means that N does not have to get very large, before the average value of each free parameter is unfathomably large.

All right, so with these hypotheses, we likely live in an unfathomably big universe, so what? Well, here is the problem. In the calculation so far, we have only summed over "sentences" of one particular template. But for the full calculation of the prior probabilities of each experience, we need to sum over the full set of all "sentences" with N or fewer "words." And the smaller the template is, i.e. in terms of a, the larger the average of each free parameter will be. Since the growth of this average will be incredibly fast with respect to $N - a$, and since the number of total observers in a universe will be proportional to e.g. the extent of space, we will therefore get that virtually all observers will be located the universes where a is is minimal (in terms of still being able to support some sort of conscious life).

With this hypothesis, we will therefore with all likelihood live in a universe whose laws, when disregarding the free parameters, are described by the the exact minimum amount of information possible in the LoE, while still being able to support conscious life. However, this does not immediately seem consistent with the low-entropy universe that we observe around us. Surely a universe filled evenly with gas would be a simpler universe to describe, would it not? And since it will still happen from time to time in such a universe that 'brains in a vat' will assemble out of pure coincidence, as we have discussed above, such a universe will still contain *some* conscious observers, even if they are far between in time and space. Therefore, we also reach the unsatisfactory 'brain in a vat' hypothesis this way.

We cannot rule out this theory completely, however, since we do not know how the "language" of the LoE is constructed. And as Schmidhuber [4, 5] points out, there are very simple algorithms who might still support conscious life. These are algorithms that basically simulate all possible programs (including their own) by simulating every ith step of every nth program in some order. Since these algorithms generate all the programs that they simulate at runtime, they can be written by relatively short programs themselves. Therefore, if the simplest life-supporting "sentences" of the LoE is equivalent to such an algorithm, the hypothesis in question might not lead to complete chaos, for all we know.

8 The multiverse as a consequence of the Logic of Everything deducing itself

Even though the hypothesis that we live in a universe of minimal information might be possible, it is worth considering other potential theories. And it just so happens that there is another theory which not only orders all the observers of the multiverse in a way which does not seem to lead to complete chaos, but in fact, it also gives a more well-founded reason for *why* this order should be used when calculating the prior probabilities of each experience.

We have discussed how it might be the case that the LoE is more abstract and complicated than anything which we can comprehend, yet it might still contain something which can be thought of as a "language" of existence. We will now go even further and propose that the LoE also includes something which can be thought of as a "deductive system." This then leads to a

[11] See e.g. Cranmer [7].

natural hypothesis: What if the multiverse is simply a consequence of the LoE deducing truths about itself?

This hypothesis means that at the beginning of everything, there was just the fundamental truths of existence, serving as the axioms of the "deductive system," as well as the fundamental set of "words" from which new meaningful "sentences" can be constructed. And from this starting point, the LoE then deduces all truths about existence, following what we can think of as some fundamental "rules of inference."

We can then further hypothesize that while the LoE deduces the infinite set of all such truths, it also "understands" these truths, and that this "understanding" for all intents and purposes brings the things that the truths concern into existence. And in particular, we can hypothesize that whenever the LoE "learns about" and "understands" some conscious experience, it also "feels" this experience. Thus, with these hypotheses, we might therefore all essentially just be the "thoughts" of the LoE, in a sense.

Now, since we have already personified this 'Logic of Everything' quite a bit by using terms such as "understands," "feels," and "thoughts," a natural critique of this hypothesis is therefore that it is simply a reinvention of a Supreme Being, a God of all existence. And indeed, if someone's beliefs in a God are consistent with a personification of this self-deducing LoE, then those beliefs will indeed be a model of this theory. However, since the theory of a self-deducing LoE does not need for this personification to be taken literally, it is broader than a theory of a Supreme Being.

In particular, most beliefs in a Supreme Being also assigns a separate consciousness to this being itself. But with the theory of a self-deducing LoE, it is also natural to hypothesize that the *only* instances in which the "awareness" of the LoE leads to actual consciousness is when it becomes "aware of" and "understands" conscious experiences, such as ours. In a sense, one might say that while the LoE might be "aware," it is not necessarily *self*-aware. It might be. But it also might not be.

9 Some nice qualities of the theory of a self-deducing Logic of Everything

This hypothesis of a self-deducing LoE first of all has the nice quality that it can explain something which is otherwise quite hard to explain: *time*. For if everything exists in equal measure, surely it would mean that all times exist simultaneously, and that 'time' is only a meaningful concept from the subjective point of view of an observer. Now, this does not sound completely unreasonable, even if it is a bit strange to think about. But with the hypothesis of a self-deducing LoE, we do not need to accept this strangeness, since the global time of the multiverse will then simply be given by whatever "time step" the LoE has reached in its infinite "computation" of everything that is true.

Furthermore, the hypothesis also solves another potential objection that one might have against multiverse theories in general, which is that they often hypothesize that an infinite number of realities exist simultaneously at any time. This is not the case for the hypothesis of a self-deducing LoE. In some models of the theory, there will be an infinite number of realities happening at the same time, but in others, there will not. In particular, if one assumes that the "time" it takes for the LoE to formulate a "sentence," perhaps by constructing it one "word" at a time, also counts as a time step in the "computation," then the LoE will only be in the process of deducing a finite number of truths at any given time! And in fact, if one also assumes that there exists a total order of all the possible deductions of the LoE, and not just a partial one, there might even be only *one* thing happening at any given time in the multiverse. (And at the moment that you are reading or hearing this, the thing that is currently happening is you.)

The fact that each truth deduced by the LoE could have what we can think of as a "computation time" also means that the theory might avoid the problem of 'complete chaos.' Looking back at the arguments of Section 7, we thus see that even if we reached the same conclusion that the most simple universes would be much larger on average (for any given N) and include many more observers than all other universes put together, the time it would take to "compute" those large universes might simply be inversely proportional to their size. And in this case, the frequency of the truths being deduced about observers living in a large universe, or a long-lived one for that matter, would not be greater, after all, than that of observers living in smaller universes.[12]

If the effective frequency of the observers is proportional to the "computation time" of the given universe, however, we might get the prediction, as pointed out by Schmidhuber [5], that the multiverse has a *resource-oriented bias*, where physical laws that are relatively easy to compute is favored over physical laws that are hard to compute. But for the hypothesis of a self-deducing LoE, one *could* hypothesize, that the "time" it takes the LoE to deduce truths about the dynamics of physical particles is generally insignificant, perhaps even infinitesimal, compared to the "time" it takes to deduce and "understand" a given conscious experience. Therefore it might not significantly matter for the prior probabilities of the observers whether they live e.g. in a hard-to-compute quantum mechanical universe or a relatively easy-to-compute classical or semiclassical universe. This is certainly a nice quality since we do seem to live in a quite hard-to-compute, quantum mechanical universe.

And last but not least, the hypothesis of a self-deducing LoE might also finally yield a potential solution to the whole 'chicken or the egg' problem of existence. The trouble is that whatever we have at the root of all existence, for instance a Supreme Being, a Great Programmer (as hypothesized by Schmidhuber [4, 5]), or a collection of mathematical structures (as hypothesized by Tegmark [1–3]), we are then typically obliged to ask ourselves: What process brought this central thing into existence in the first place? What is the underlying logic that determined that it was exactly *this* thing which became the center of all existence, and not one of the other possibilities? But with the hypothesis of a self-deducing LoE, the answer would simply be: The underlying logic behind all of existence *is* all of existence.

10 Ethics when consciousness springs out from the very center of all existence

Another nice consequence of having a multiverse like that of the self-deducing LoE, or indeed of any other theory where the number of universes is infinite, and where consciousness springs out from the very center of all existence,[13] is that philosophical egoism actually ends up yielding the same conclusions as utilitarianism.

The reasoning for this is as follows. If the multiverse lives on forever, continuously containing parts of itself with observers that are alive and conscious, then all possible lives will be lived

[12] By the way, the same thing can be said about universes that obeys the principle of the many-worlds interpretation of quantum mechanics versus universes where the wave function instead collapses from time to time (perhaps when some great enough level of decoherence is reached). The many-worlds universes will simply be all the more hard to compute, and therefore they will *not* be exponentially more likely, from the observers' point of view, than the wave function-collapsing universes.

[13] That 'consciousness springs out from the very center of all existence' can be said about many other theories as well, including all that assume the principle of materialism, as well as some deistic theories, namely ones where 'God is everything,' and where everyone is therefore just a part of God.

by its collective observers an infinite number of times. Some lives will be far more likely than others (otherwise we would have 'complete chaos'), but they will all be lived from time to time, even if they require a great number of unlikely coincidences. One might think of the space of all possible lives as a landscape of hills and valleys, mountains and canyons, where the more likely lives are paths through the landscape that only follow the valleys and canyons, and where the unlikely lives are the paths that cross one or several hills or mountains. If one then compares any two given paths, there will always exist a continuum of paths in between them, not just for any two paths that start at the same point, but even for paths that start at completely different points. And even if the lives that some of those paths represent are incredibly unlikely in the multiverse, these lives will still be lived from time to time, albeit very infrequently.

This means that not only will an observer with exactly the same life as you live again and again an infinite number of times in the multiverse, but so will observers that live some lives that are only some small variations away from yours. Furthermore, the lives that are small variations further away from those variations will also be lived again and again, etc.

Now, combine this fact with the hypothesis that the source of your consciousness springs out from very center of existence, such as in the hypothesis of the self-deducing LoE, where all conscious experiences are essentially "lived" by the LoE itself. Then you get that any observer in the future who lives a life which is just a very small variation away from yours will essentially *be* you for all intents and purposes. And if that person is you for all intents and purposes, then so will the next person be that lives a life which is just one small variation further away from that. Therefore, if this logic holds up, we then get that, with variations upon variations, you will at some point in the future live every single possible live that can be obtained by starting with your current one and then repeatedly making small variations to it. And since all possible lives in the universe can be reached this way, even if you have to cross "mountains" to get there, we thus get that you will live all possible lives in the universe, again and again.[14]

This result has the happy consequence that it all of a sudden makes the question of what defines ethically right or wrong actions on the fundamental level very clear-cut, even if you are a philosophical egoist. For even people of that persuasion would then still have to agree to the following principle of ethics.

Principle. *Live your life as if you know that you are going to live all possible lives in your universe, including the lives of the people and other beings whom your actions affect.*

It might be noted, that while the obvious consequence of this principle is that you should not prioritize your own happiness and satisfaction over others' on a fundamental level,[15] it is also important to remember that the reverse is true as well. One should not sacrifice oneself *too much* for others, i.e. not to the extent that resulting pain and dissatisfaction ends up lowering the total. For it is not just *you* who have to live *their* lives at some point in the future. *They*

[14] This prospect might sound frightening to some, but remember that you will not take any memories with you to the next life, at least not according to this hypothesis alone. In fact, the number of lives that you have already lived in the past will be infinite on average whenever you are reading or hearing this. And sure, there has been a lot of pain in the past history of the Earth, but if one picked out any person at random from the past and asked if they regret having been born, they will probably answer 'no' most likely. Furthermore, if only our civilization survives all nuclear threats, and not least survives the coming singularity of AI, the happiness of all future lives, when we multiply and extend ourselves out in space, will probably outweigh all the pains of our past and present by far.

[15] Of course in reality, it still often makes sense for a person to *focus* more on their own happiness than all others', first of all because they will most often be the person best fit for that that job, and second, because going *too much* against one's own natural instincts and desires might lead a person to break in the end, leaving all relevant parties worse off. (It also goes without saying that it generally makes sense for a person to prioritize the people that are close to them.)

will also have to live *your* life.[16]

11 Conclusion

We argued that existence, and in particular 'conscious experiences,' might require more than simply mathematics to explain, and that one might therefore want to extend the Mathematical Universe Hypothesis (MUH). This lead us to propose a hypothesis that each universe defines its own 'consciousness laws,' similarly to its physical laws. We then argued that the original MUH seems to run into a problem of 'complete chaos,' and went on to investigate if there is a way for other theories to avoid the same problem. This lead us, in the end, to propose a hypothesis of a self-deducing Logic of Everything, which does not seem to necessarily lead to complete chaos, and which also has several other nice qualities to it. And finally, we discussed some existential consequences of such a theory of existence, and other multiverse theories like it, including what fundamental principles of ethics can be derived from it.

References

[1] M. Tegmark, *Is "the theory of everything" merely the ultimate ensemble theory?*, arXiv:gr-qc/9704009.

[2] M. Tegmark, *Parallel Universes*, arXiv:astro-ph/0302131.

[3] M. Tegmark, *The Mathematical Universe*, arXiv:0704.0646 [gr-qc].

[4] J. Schmidhuber, *A Computer Scientist's View of Life, the Universe, and Everything*, arXiv:quant-ph/9904050.

[5] J. Schmidhuber, *Algorithmic Theories of Everything*, arXiv:quant-ph/0011122.

[6] D. N. Page, *Is Our Universe Decaying at an Astronomical Rate?*, arXiv:hep-th/0612137.

[7] S. R. Cranmer, *Intuitive Explanations of Infinite Numbers for Non-Specialists*, arXiv:2401.07346 [math.HO].

[8] P. J. Steinhardt. (2014). Theories of Anything. Edge.
https://www.edge.org/response-detail/25405

[9] P. Davies, *The Goldilocks Enigma*, (Penguin Books, London, 2007).

[10] D. J. Griffiths, *Introduction to Quantum Mechanics*, 2nd ed. (Pearson Education, Edinburgh, 2014).

[11] R. Shankar, *Principles of Quantum Mechanics*, 2nd ed. (Springer, New York, 1994).

[12] E. S. Abers, *Quantum Mechanics* (Pearson Education, London, 2004).

[13] P. Hedegaard, *Statistisk Fysik*, 2015 version, Lecture notes, Niels Bohr Institute, University of Copenhagen, 2015.

[16] Similarly, do not kick yourself too much and bring yourself too much down over past mistakes; only to the extend that you feel that you have learned whatever lesson there was to learn. Anything above that is not constructive. For remember that the people that you have wronged or caused discomfort will also experience making the same mistakes when they live your life at some point in the future of the multiverse. Therefore, do not (generally speaking) make their pain double by making yourself feel an equal (or greater) amount of pain, *if* you are able to avoid it.