Let's Hope We're Not Living in a Simulation

Eric Schwitzgebel
Department of Philosophy
University of California, Riverside
Riverside CA 92521
USA

April 1, 2024

Let’s Hope We’re Not Living in a Simulation

Abstract: In *Reality+*, David Chalmers suggests that it wouldn’t be too bad if we lived in a computer simulation. I argue on the contrary that if we live in a simulation, we ought to attach a significant conditional credence to its being a small or brief simulation. Our existence and the existence of many of the people and things we care about would then unfortunately depend on contingencies difficult to assess and beyond our control. Furthermore, all the badness of the world would appear to reflect the gods’ intentional cruelty or callous disregard. A large, stable rock is a more dependable and less axiologically troubling fundamental ground for reality.

Keywords: Chalmers, David; simulation hypothesis; problem of evil; skepticism; theodicy; virtual reality

Word Count: 3497 (including references)

Let’s Hope We’re Not Living in a Simulation

1. The Simulation Hypothesis.

I assume the reader is familiar with the *simulation hypothesis* – the idea that we are artificial intelligences living in a virtual reality. In *Reality+*, David Chalmers argues that the chance we live in a simulated reality is “at least 25 percent or so” (Chalmers 2022, p. 101; cf. Bostrom 2003). I have a substantially lower credence, about 0.1% to 1% (Schwitzgebel 2024, ch. 4) – small, but high enough to be concerned about the possible implications.

In this commentary, I’ll argue that we should very much hope we’re not living in a simulation. My worry is this: If we live in a simulation, we have excellent reason to doubt that there will be a tomorrow. We exist, perhaps briefly, at the whim of powerful entities whose kindness we have little reason to expect.

2. The Size Question.

Consider a large virtual reality: the entire observable universe – all 93 billion light-years of it, with every star and planet simulated in microscopic detail throughout hundreds of billions of galaxies – a reality enduring many billions of years. Consider a small virtual reality: you, your immediately perceivable environment, and nothing else, in a simulation that started five minutes ago and which will last another five minutes before deletion.

Between these extremes lie many possibilities, some small enough to be epistemically catastrophic – small enough that a substantial proportion of our everyday beliefs would be false or lack reference in virtue of the nonexistence of things or events whose existence we ordinarily take for granted. An epistemically catastrophic virtual reality might be geographically small, for example if only one city exists; or it might be temporally small, for example if it was created ten years ago; or it might have a small population, for example if most of the seeming-people are mock-up sprites without real conscious experiences (Helton 2021). We normally assume that we had childhoods and live on a planet with billions of people. If such assumptions are false, we are wrong about many things of importance.

How confident ought we to be that if we inhabit a virtual reality the reality is large enough to be epistemically non-catastrophic – that the world contains more or less all of the things we care about, plus a reasonably deep past, plus a reasonably long future, and billions of people? Call this the *Size Question*. An optimist about the Size Question holds that we ought to be confident that if we are sims, we don't live in a catastrophically small simulation. The pessimist denies this. I endorse pessimism. Whatever credence we attach to the simulation hypothesis, we ought to attach a substantial conditional credence (10%? 50%? 90%?) to catastrophic smallness.

How should we assess whether optimists or pessimists are closer to right about the Size Question? I see three families of approach.

One approach professes radical ignorance. We have essentially no idea how to evaluate the Size Question. We have no idea what would motivate the creation of an Earth-like simulation; or how many such simulations exist; or how expensive they are to create; or what hazards they are subject to; or their typical size, duration, or features; or how base-level reality generates our virtual reality. Such radical epistemic humility is not an unreasonable stance. However, it appears to justify pessimism rather than optimism. Radical ignorance of this sort is incompatible with the optimist's confidence.

A second approach appeals to Moorean certainty or Wittgensteinian "hinge epistemology". On this approach, we are justified in treating some select propositions as fixed points in our reasoning. Among these propositions might be “here is a hand” and “billions of people exist”. This approach, strictly implemented, can conveniently solve any skeptical problem. But it should not be strictly implemented. Recent advocates of hinge epistemology, such as Coliva (2015), acknowledge that hinges (e.g., “no one has ever walked on the Moon”) can be revised under the right conditions. The hypothetical discovery that you’re a sim is exactly the sort of situation that should trigger reconsideration of one’s hinges. To insist that you know, or can be certain, that if you are living in a simulation it is a large one, without further supporting grounds, simply begs the question. It is not a justifiable defense of optimism.

A third approach, which Chalmers appears to favor, and which I also favor, attempts to assess the evidence that we occupy a large or a small simulation. Chalmers considers small simulations of two types: local and temporary.

Might the simulation contain only one city (Chalmers 2022, p. 442-444)? Stipulate that the city has existed for at least a hundred years, but nothing beyond it exists. Everyone in the city exists, and we have real conversations with each other. The room you are in exists, and the building, and the roads – but everything stops at the city edge. Anyone looking beyond the edge sees, presumably, some false screen. If they travel past the edge, they disappear from existence; and when they return, they pop back into existence with false memories of having been elsewhere. News from afar is all fake. Unless you grew up in the same city, your childhood is fake.

Addressing this possibility, Chalmers suggests that “the most obvious objection” to the local simulation is that it lacks simplicity. It lacks simplicity, presumably, because the computer will somehow need to coordinate everyone’s false memories, generate appearances of fields, buildings, and roads beyond the city’s boundaries, create the fake news, and so on, without noticeable contradiction or inconsistency. In contrast, Chalmers says, “global simulations just require simulating a few simple laws of nature and letting the simulation unfold” (p. 444).

I doubt we should be so quickly satisfied with a simplicity response. Appeals to simplicity are often indecisive. The world might not be so simple; or it might be simple in some respects but not others. Also, it’s unclear whether a planet-sized simulation is actually simpler than a city-sized one. A planet-sized simulation will have many more objects, many more people, and many complex, distant events. Simplicity of law is only one dimension of simplicity.

Chalmers acknowledges that large simulations are likely to be much costlier than local simulations. Simulators might for this reason, he suggests, create only Earth instead of a whole galaxy of stars, if they are mainly interested in our lives (p. 444). Everything beyond the Solar System could easily enough be mock-up patterns of light, misleadingly designed to resemble a universe of stars and galaxies beyond. The creators might design a simulation bounded at the edge of the Solar System, even if that means not just choosing a few simple cosmic laws and letting everything run forward from the beginning.

But analogous reasoning applies to the city. Maybe one city is all our creators want or need. It might be easy enough to create fake boundaries, fake news, and fake memories, all nicely coordinated. The simulation needn’t be perfect. If the city’s inhabitants start to notice inconsistencies, maybe the inconsistencies can be repaired post-hoc and inhabitants’ memories rewritten. If our creators want a solo city, they might well have the resources to fool us well enough. We’re really in no position to confidently assess whether it’s easier and more efficient to create a whole planet or a whole observable universe for the sake of a city or whether it’s easier and more efficient just to create the one city and somehow deal with the problem of faking the world beyond. Simplicity arguments cannot justify high confidence that if we live in a simulation, it is planet-sized rather than only city-sized.

How about a temporary simulation, one created on April 1, 2024, with memories, historical documents, old buildings, fossils, etc., already in place? This simulation would be catastrophically small: All of our historical beliefs would be false, we would never have been children, and most of the people we regard as deceased would never have existed. Chalmers also argues against the temporary simulation scenario on grounds of simplicity. He suggests that “the obvious way” to create the right kind of fossil records and so forth would be to run a detailed simulation of the past – in which case we’re not in a temporary simulation after all (p. 446).

Again, however, considerations of cost and simplicity might compete. If the simulators are only interested in us as we exist now, they might not want to pay the cost of running a full simulated Earth for millions or billions of years, and they might be able to avoid that cost by generating a plausible enough distribution of historical records, memories, etc. Again, I doubt we can confidently assess how the cost of a long-enduring sim balances against the difficulty of seeding a well-organized start date.

As Chalmers notes, some temporary simulations might be easy to create – for example, one person awakening from a nap in a dark room (p. 446). If the person doesn’t survive long, they won’t have much chance to check for flaws and shortcuts. Another simple model might be a few dozen people together in a room listening to a philosophy talk. Do we need a whole history and fossil record in place? Do we need North America to exist? No one will really be checking; and even if a few of us fire up our phones for sports news or whatever, plausible inputs could probably be faked for the several minutes we are here existing together.

Chalmers (see also Bostrom 2011) observes that even if large simulations are less common than small ones, large simulations will contain many more people, with the possible consequence that any individual person is likely to be in a large simulation (p. 139-140). Suppose, for example, that the relevant technological society creates a million simulations of a single person alone in a room and only one simulation of a planet of a billion people. On a plausible self-location principle, it's then a thousand times likelier that you are in the large than in the small simulation, since that's where most of the 1,001,000,000 people who exist are. I accept that this argument justifies not partitioning your conditional credence evenly among all possible simulations (though for some concerns, see Lewis and Fallis 2023). But again, cost considerations might loom large. Consider our own early 21st century simulations. Rarely do we simulate worlds with billions of individually modeled inhabitants. We run lots of small simulated scenarios as games and scientific projects. Our planet-scale simulations, such as climate models, model people only as aggregates. While the simulators of our world *might* be unconstrained by resource limitations or legally required to create only simulations that are not epistemically catastrophic, it seems unreasonable to have high confidence that this is the case.

I find myself somewhat unsure how to map Chalmers' view onto pessimism versus optimism regarding the Size Question. In general, Chalmers acknowledges sources of skeptical doubt, admitting uncertainty. On the other hand, he argues that considerations of simplicity tend to speak against various skeptical scenarios and that there's reason to expect conformity between the structures of our experience and the structures of the world. In any case, I want to be clear on this point. If we adopt an evidential approach to the Size Question, the evidence does not decisively favor a positive answer. We can mine Chalmers' remarks for some considerations in favor of thinking that the simulation would not be catastrophically small, but those considerations – mainly the appeal to simplicity and a self-location principle – don't yield decisive results and are opposed by the general observation that, to the extent we can guess about such things at all, it seems reasonable to suppose that large simulations will be higher cost.

The most epistemically reasonable position is doubt. Whatever credence you assign to the simulation hypothesis, you should assign a substantial subportion of that credence to the possibility that you, or we, live in a small simulation in which much of what you take for granted about the world is false.

3. The Pathetic, Cruel, or Indifferent Gods.

Stipulate that we know, somehow, that we live in a large, stable virtual reality. Planet Earth exists, has existed for a long time, and will exist long into the future, with billions of people leading lives of approximately the sort they think they're leading, even if the fundamental metaphysics is surprising. Would this be bad? Would this be worse than living in the base level of reality? Chalmers suggests that in some respects it might be a little worse (Chapter 17). For example, the world would lack the full, rich history that we normally assume it has, and maybe that history matters to us. However, if the simulation is complete enough – if ordinary things really are more or less how we think they are – then our reality has most of what gives life value: real emotional states, real interactions with other people, real ethical decisions, real accomplishments, real engagement with interesting and challenging environments. So there's not much reason, Chalmers argues, to hope that we aren't living in a simulation, as long as we know that it's appropriately large and stable (though see Avnur 2023 for an argument that even a large, stable simulation would be epistemically catastrophic).

However, despite having a chapter on theology (Chapter 7), Chalmers does not, I think, sufficiently explore the unfortunate theological consequences of the simulation hypothesis. We have ethical or axiological grounds for hoping we aren't sims, in addition to the epistemic and prudential grounds already discussed.

A simulation presumably requires a simulator. (If a mindless swamp plant unthinkingly belched forth our world as a side effect of some digestive process, that's not the simulation hypothesis.) Our simulator, or others of its kind, or others not quite of its kind, presumably also designed our reality. Perhaps this simulator has the power to delete our reality or to interfere with it, creating "miracles" that violate what we regard as the ordinary laws of nature. This simulator exists outside of our spatial manifold, uncontained by our spatial dimensions, presumably capable of existing even if our whole reality ceases. The simulator's time might differ radically from our time. Maybe the simulator can pause us without our realizing it. Maybe the simulator can rewind us to a save point, tweak a few things and then restart, in a sense changing the past. Maybe the simulator can copy our whole world. Maybe they can change the laws of nature. All of this might be true even if at the base level of reality, the simulator is unimpressive – some foolish adolescent gamer living in their parents' basement.

If Zeus is a god, then our simulator is a god. Even just the power to choose to launch or not launch our world, combined with existence outside of our spatial manifold, is arguably sufficient for the simulator to be a proper referent of the term "god". At first, Chalmers seems to agree (p. 128), though later he demurs on grounds that he would not want to worship the simulator (p. 144). I doubt worship-worthiness is a necessary criterion of godhead, so I'll continue to call the simulator a god, but I take this to be a terminological point. "God" turns out, perhaps surprisingly, to be a relational term: An entity can be a god relative to one reality and a non-god relative to another reality (see also Schwitzgebel 2019, ch. 21).

We can now do some *natural theology*. That is, we can examine the world around us with the aim of making educated guesses about the properties of God or the gods. One striking fact regularly noticed by natural theologians is this: The world contains plenty of apparently unnecessary suffering (e.g., tooth decay) and moral evil (e.g., the Holocaust). This appears to force a theological choice: Either God *cannot* prevent all the bad things that happen, or God *prefers not* to prevent all the bad things that happen.

Consider *cannot* first. God closed her/his/its/their seventeen bug-faceted eyes and pressed the launch button, hoping for the best. "Go, little world!" Then God let things alone, knowing intervention isn't possible, maybe toodling off to other tasks, never looking back; or maybe God watches in impotent horror as Genghis Kahn, Hitler, and Stalin kill their millions, as children die of painful cancer, as plagues and starvation ravage us, as billions of people suffer badly designed bodies prone to pointless sinus infections and back trouble. Maybe God tries to help but proves ineffectual. Call this the *pathetic God* possibility. God couldn't even have given Hitler a heart attack? That minor, virtually undetectable action, which would have struck nobody as distractingly miraculous, could plausibly have prevented enormous atrocities. Our creator has so little control over their creation? That's sad, pitiful – a terrible design mistake. A world with billions of genuinely conscious, suffering people needs a good user interface.

More likely – to the extent we can assess likelihood – if the gods have the power to create a simulated world, they have the power to make adjustments. They just prefer not to. Maybe God is an angry adolescent who delights in our cute little guns and military costumes. Maybe God is a scientist who sees us as lab animals whose suffering is irrelevant as long as the hypotheses of interest can be tested. Maybe God is an artist whose audience is awed by the cruel display. "Don't you dare touch Hitler!" God says to the leering viewers. "The perfection of my artistic vision requires the Holocaust." The audience departs in tears, and God racks up the attendance fees and grant money.

Of course, a rich theological tradition suggests responses. Following Leibniz (1710/1952), we might posit that this is the best of all possible worlds, or at least the best of all technologically feasible worlds. No simulator could have made things better. Give Hitler a heart attack and something even worse than the Holocaust would have happened. God dare not save that innocent three-year-old from the slow, painful death of Tay-Sachs disease, by curing the disease or at least giving a quicker death, otherwise… well, something. Following Stump (2010), maybe God presents every person only the suffering required to give them the best chance to flourish: Every child who starves to death is suffering because starvation unto death is their best chance to achieve the desires of their heart.

It is perhaps not strictly impossible that Leibniz or Stump are correct. As Hume (1779/1947) emphasizes, if one knows in advance for certain that an omnipotent, benevolent God exists, one might justifiably accept a view of this sort. But an even-handed natural theology, which examines the world empirically to discover the features of God without antecedently assuming God's goodness, hardly suggests that this is the best of all possible worlds or that every human agony serves a great soul-building purpose.

To the extent we can evaluate likelihood, the power to design and create suggests the power to intervene and improve. Even casual examination suggests fixes that should be easy for civilizations of such immense power. The absence of such fixes suggests that our creators are cruel or at best grossly irresponsible.

If so, our world is ethically and axiologically worse than if evil and suffering arise from the indifferent forces of nature. It is worse in the same way that it's worse for a child to die due to murder or neglect than for that same child to die by unforeseeable accident. Because our creators are responsible for our existence and for our relatively happy or miserable state, they owe us benevolence. To all appearances, they are abhorrently unbenevolent. What kind of ethics review board approves of *this* experiment?

4. Therefore, a Large, Stable Rock Would Be Much Better.

Let's hope we're not living in a simulation. If we are sims, our existence and the existence of many of the people and things we care about depend on contingencies difficult to assess and beyond our control, and all the badness of the world appears to reflect the gods' intentional cruelty or callous disregard. A large, stable rock is a more dependable and less axiologically troubling fundamental ground for reality.

References

Avnur, Yuval (2023). *Reality+: Virtual Worlds and the Problems of Philosophy, 98,* 107-111.

Bostrom, Nick (2003). Are we living in a computer simulation? *Philosophical Quarterly, 53,* 243-255.

Bostrom, Nick (2011). Bostrom's response to my discussion of the simulation argument. Blog post at *The Splintered Mind* (Sep. 2).

Chalmers, David J. (2022). *Reality+.* W. W. Norton.

Coliva, Annalisa (2015). *Extended rationality.* Palgrave.

Helton, Grace (2021). Epistemological solipsism as a route to external world skepticism. *Philosophical Perspectives, 35,* 229-250.

Hume, David (1779/1947). *Dialogues concerning natural religion,* ed. N.K. Smith. Bobbs-Merrill.

Leibniz, G.W. (1710/1952). *Theodicy,* ed. A. Farrar, trans. E.M. Huggard. Routledge & Kegan Paul.

Lewis, Peter J., and Don Fallis (2023). Simulation and self-location. *Synthese, 202* (180).

Schwitzgebel, Eric (2019). *A theory of jerks and other philosophical misadventures.* MIT Press.

Schwitzgebel, Eric (2024). *The weirdness of the world.* Princeton University Press.

Stump, Eleonore (2010). *Wandering in darkness.* Oxford University Press.