

A REFUTATION OF THE SIMULATION ARGUMENT

© Dan Bruiger, 2023

1. Introduction

The Simulation Hypothesis is the idea that we are actually living in a simulation rather than physical reality. The Simulation Argument reasons that this is probably so. The roots of this notion are ancient, going back at least to Plato's Cave. The Vedas had taught that the appearance of reality is illusory, and Chuang Tse had wondered whether he was a man dreaming of being a butterfly or a butterfly dreaming of being a man. Indeed, the parallel with sleep is the idea that we are actually living in a dream even in our apparently waking state. In modern times, such skepticism is grounded in Descartes' argument against the reliability of the senses: one's consciousness could be faked by an "evil genius" who had control over the nervous system so that one would never be the wiser.¹ This is the origin of the idea of the "brain in a vat," a thought experiment explored in philosophy and in science fiction, such as in the *Matrix* films.

The intuition that phenomenal experience could be illusory is probably as old as self-consciousness itself. In modern times, it follows upon the understanding that the contents of consciousness are produced in the brain and nervous system. It is given further impetus by the computational theory of mind, artificial intelligence, and virtual reality. The latter extends to the contemporary gaming scene, whose seductions can be compelling, especially for a cadre of youth who spend a major portion of their waking life online, attending to simulated environments rather than real ones. Some people seem to have seriously wondered whether they were actually living in a computer simulation. And for some, their doubt has had serious real consequences.²

The computer technology of simulation has aggravated, updated, and made tangible an ancient anxiety. It is therefore relevant to critically examine the seminal paper that articulated and formalized this anxiety in modern terms: Nick Bostrom's "Are You Living in a Computer Simulation," published in 2003 in the *Philosophical Quarterly*.³ I will examine the assumptions and logic in that paper, and show how the question itself is loaded.

2. Dubious assumptions

The paper begins by explicitly taking as given the modern concept of 'substrate independence': that *mind* or *consciousness* is not a function of a material substrate but of organization and structure only, which (in the computational metaphor of mind) amounts to computation. Accordingly, in principle, mind can be simulated by computers, whose substrate is non-biological and intentionally engineered. How well or closely a brain or its mind can be thus simulated is held to be a matter of computational power only. The computational theory of mind, like artificial intelligence generally, tacitly assumes that intelligence can be divorced from its natural occurrences, which are living organisms. The free-standing computer is disembodied in two senses: its relation to the physical world is mediated by human beings; and the computer is not an autonomous agent acting on its own behalf for its own purposes.⁴

Whatever the possibilities of disembodied intelligence,⁵ the possibility of disembodied *consciousness* is the overriding presumption that makes it meaningful even to raise the question

¹ The idea and ideal of omniscience (or, as we now say, unlimited computational resources) is also ancient. Descartes' "demon" was the model for Laplace's Demon and Maxwell's Demon.

² For example, in the 2021 documentary *A Glitch in the Matrix*, a teenager describes how he shot to death his mother and father to test whether he was living in reality or a simulation.

³ <https://www.simulation-argument.com/simulation.pdf>

⁴ That is, it is not an autopoietic system.

⁵ See my paper, "What is Intelligence in the Context of AGI?" [<https://stanceofunknowing.com/collected-writings/>]

posed in the title of Bostrom's paper. We must first ask, what can it *mean* to be living in a simulation? The title addresses 'you' as a living and experiencing subject. You may naturally believe that you are (or "have") a real body, on which your consciousness depends. Yet, like Descartes' skepticism, the question puts this belief in doubt. You could merely *think* you have a real body, alive in a real world. However, for you to be so deluded presumes that consciousness *can* exist without embodiment. This is the tacit assumption underlying the question 'are you living in a simulation?'

Then comes the second major assumption: a "post-human" society will create ever more powerful computers with which to do simulations. Bostrom points out that while "simulating the entire universe down to the quantum level is obviously infeasible,...much less is needed – only whatever is required to ensure that the simulated humans, interacting in normal human ways with their simulated environment, don't notice any irregularities." Why would it matter to simulated people that they live in a faked world? Obviously because real people are (for whatever reasons) concerned with the truth of their situation. And real people are capable of interacting in other than "normal human ways" with their environment. That is the whole purpose of science, whose instrumentation can be more sensitive than the natural senses and more powerful than the native mind. After all, we are currently trying to develop "superintelligence" to aid in making discriminations beyond our natural capacities. We might rightfully insist that the *question* at least be scientific. That is, there should be a feasible test or experiment to determine whether we live in a simulation. But as the computational power to question appearances grows, so would the power required to maintain the appearances scientists are trying to penetrate. Can there be a resolution to that contest?

This brings us to the third major assumption: that *simulated* humans could be conscious cognitive agents in their own right. Could simulated minds be motivated and capable of detecting irregularities and unmasking deceptions? While their attempts at detection might be described behaviorally from the outside, presumably they would be conscious like natural humans and have their own phenomenality. (This is necessary to even pose the dilemma, in which we are the conscious humans wondering whether they are simulated.) Yet, by definition, the characters in a virtual reality are *not* "agents" in the sense we deem ourselves and other creatures to be. Rather, they are *patients* in the archaic sense of simply being the passive victims of actions performed by the genuine agents who create the simulation. For, as Bostrom notes, "Should any error occur, the director could easily edit the states of any brains that have become aware of an anomaly before it spoils the simulation."

Here we might pause to wonder what it means to "spoil" the simulation. To simulate our existence could simply mean to re-create the natural human situation, in which some people do in fact doubt the verity of their perceptions or the reality of the world. If human experience is indeed a simulation, their suspicion is simply part of the simulation and in no way destroys it. The language used ("spoiling") points to the *intent to deceive*. An error or glitch in the program could reveal the deception involved, showing that there must be a "true" reality behind the false appearances. It is the deception that is spoiled, the intent that is foiled. This was in fact the worry of Descartes' thought experiment: the wariness of being deliberately deceived by a mischievous agent. Descartes' official (and evasive) solution was that God in his benevolence would not permit such maliciousness. The wily Descartes did not expound on the possibility that the existence of God might be part of the deception!

What if *we* are the agents creating the simulation? Why would we want to mischievously foist such delusion on unsuspecting artificial subjects? The goal of actual VR is a satisfying entertainment, a convincing substitute for reality, but not a deception. The real live user lives in a primary reality and knows the difference. So, we are considering two distinct situations: verifiably real observers deliberately entering a VR experience; or putatively real observers finding themselves in a 'world' that may or may not be a simulation. The simulation hypothesis pertains to the latter, so what is really in question is whether the putative observer is real. What test is there for

that? Pinch oneself? Subjects who find themselves uncertain about the reality of their world must also doubt their own reality. They either have no knowledge/memory of living in a different (primary and real) world or else they suspect that such memory is faked. But what can “fake” mean without a genuine version?

The next assumption is that, with enough computational power, it should be feasible to make simulated environments that fool simulated people. But, even if the latter are possible, can you fool all the people all the time, especially if they are making efforts to find the truth through augmenting their own (simulated) computational power? If the creators of the simulation have unlimited computational power to deceive, why couldn’t the simulated inhabitants of the simulation (assuming they are conscious and intelligent) develop unlimited simulated computational power to unmask the deception? Bostrom suggests the main computational cost of creating realistic enough simulation “resides in simulating organic brains down to the neuronal or sub-neuronal level.”⁶ He proposes that a planet-sized computer should do the trick many times over. But, perhaps, like present real human beings, the simulated people would try to create artificial superintelligence far less gullible than the human brain. What if the people in the simulation build a simulated planet-sized computer?

Bostrom further assumes that people (whether real or simulated) would for some reason want to create “ancestor simulations.”⁷ That is, to recreate in simulated form the lives or minds of dead people. (This assumption is also essential for the Simulation Hypothesis, since we the living could be those people, resurrected in such a project.) Motivation aside, unless it were possible to exhaustively archive the brain of a living person, or their memory (which is sketchy and highly selective), there would be no sound basis to accurately resurrect a person from death. The simulacrum would be no more than a biography, an original creation based on fact. But why, granted unlimited computational resources, would future people want to replicate their ancestors or the past when they could equally well create any novel fictional characters and situations they could imagine? Why not create completely new and original simulated people if you want to create simulated peoples at all?⁸

3. Dubious logic

Bostrom now proceeds to the core of his argument, by introducing a formula that is modelled, apparently, on the famous Drake Equation to estimate the number of intelligent technological civilizations in the galaxy or universe. In this case, the goal is to estimate the “actual fraction of all observers with human-type experiences that live in simulations.”⁹ His equation has the form $n = x/(x+1)$.¹⁰ This handily transforms to $n = (1-n/x)$. As he posits, “x” is a product of three factors, two of which might well be very small, and one of which could be very large. His argument reduces to the claim that if x turns out to be very large, then $n \sim 1$ (corresponding to a probability approaching 100%). However, the smaller x is, the larger is n/x ; therefore, n could be very much smaller than

⁶ The current goal of efforts to do so, such as the Human Brain Project, is not to create conscious characters in a virtual reality but to create an artificial mouse brain that could control a real mouse.

⁷ I suspect this notion may derive from Frank Tipler’s writing, e.g. *The Physics of Immortality* (1994), which proposes that minds could be “resurrected” as simulations.

⁸ This parallels the question of why people seek immortality when nature provides fresh original minds and bodies through reproduction.

⁹ The restriction to “human-type experiences”, whatever exactly that might mean, suggests broadening the question to include any possible epistemic agent (whatever that might mean).

¹⁰ His actual formula is $f_{sim} = (f_p f_l N_l) / (f_p f_l N_l) + 1$, where f_{sim} is “the fraction of all observers with human-type experiences that live in simulations,” f_p is the fraction of all human-level technological civilizations that survive to reach a posthuman stage, f_l is the fraction of posthuman civilizations that are interested in running ancestor-simulations, and N_l is the average number of ancestor-simulations run by such interested civilizations.

one.¹¹ Like the Drake Equation, there is little basis on which to set the values of the various parameters; so, in truth, there is no way to know the fraction of observers living in simulations.

4. Dubious inferences

Nevertheless, Bostrom unpacks some consequences of *assuming* $n \sim 1$ (i.e., that it is probable that we live in a simulation). For example, he says it would trivially be wiser to bet that one is living in a simulation than not. (By the same assumption, the payoff for winning the bet would be simulated, not real.) But betting presumes a contest whose outcome is not predetermined yet can be somehow decided. There must be some criterion or test to decide the question. Bostrom doesn't discuss this. What could that test possibly be if it is assumed from the outset that it is effectively impossible for the subject to tell the difference between reality and simulation, granted the unlimited computational power behind the latter?¹² In the film *The Matrix*, telltale glitches occur in the simulation, indicating that it *is* a simulation—and also that there is a limit to the computational resources behind it.¹³

Bostrom considers that we may “have no information that enables us to predict any differences between the experiences of simulated minds and those of the original biological minds.” Apart from what ‘information’ might mean, this raises two questions: What is the fundamental difference between reality and simulation? And under what conditions is that difference detectable? The short answer to the first question is that reality is *found* and simulations are *made*. Made things (including information and knowledge) are finite and well-defined. In contrast, the natural reality that is the object of our senses and knowledge is inherently ambiguous. It might or might not be infinite in extent and complexity—we do not know.¹⁴ Human beings stand roughly midpoint in the known scale of things. At the human scale, the differences between familiar natural things and familiar artifacts seem obvious. We cannot be so certain in less familiar realms. Over centuries, we’ve extended vision with microscopes and telescopes. Bostrom proposes that post-human simulators could accommodate such an expansion of the cognitive world of simulated observers they create by filling in “sufficient detail in the simulation in the appropriate domain on an as-needed basis.”¹⁵ In fact, this is what the brain normally does!¹⁶

Since the ideal of simulation (and the purport of AI generally) is to mimic the real thing ever more closely, at some identifiable threshold the distinction between nature and artifice will evade our natural senses. On the other hand, technology extends abilities ever more. This defines a potential game of one-upmanship, a competition between code-makers and code-breakers. So, the answer to the second question depends on where exactly we stand in that game. Assuming that we *are* living in a simulation, the contest would be between our (simulated) computational resources

¹¹ An interesting recent paper considers the case where simulations are nested, arguing that the detail resolution of a given simulation would then necessarily decay with the number of nestings; the probability of living in a simulation would diminish accordingly. See: Ozzy King “Limits to the Simulation Hypothesis” preprint 2023 [https://www.researchgate.net/publication/367041320_Limits_to_the_Simulation_Hypothesis]

¹² One proposal for a test trades on an analogy between the collapse of the quantum wave function and the “moment of rendering” when new detail must be supplied to a “player” in VR. The moment of rendering is the time in the host world when detail is filled into a simulation—triggered by either by an experiment within the simulation or the demand of a conscious player. Similarly, the collapse of the wave function is either triggered by an objective event (such as in an apparatus) or by the subjective event of the observer’s awareness. See: Tom Campbell, Houman Owhadi, Joe Sauvageau, David Watkinson “On testing the simulation theory” [[arXiv:1703.00058v2{quant-ph}6 June 2017](https://arxiv.org/abs/1703.00058v2)]

¹³ The energy for which, ironically in the film, is supplied by the deluded victims themselves.

¹⁴ See my paper, “Does Nature Have a Definite Information Content?” [<https://stanceofunknowing.com/cw/Does-Nature-Have-a-Definite-Information-Content.pdf>]

¹⁵ One might as well suggest that God fills in the appearance of the world behind our backs as we turn our heads to see.

¹⁶ Our eyes see in clear detail only a tiny area of the visual field (the fovea), while we assume that same detail exists in the peripheral field. Part of the visual field is obstructed by the blind spot, in which there is no sensory stimulation, but which nevertheless appears filled in by the brain.

and the (real) computational resources of our creators.¹⁷ It seems the latter would necessarily have the upper hand.

Bostrom deduces that “if we are living in a simulation, then the cosmos that we are observing is just a tiny piece of the totality of physical existence.” He uses ‘physics’ and ‘physical existence’ ambiguously, as though there is no “real” difference between actual physical existence and simulated physical existence. This, again, is a tacit assumption upon which the simulation hypothesis depends to make any sense at all. He continues: “The physics in the universe where the computer is situated that is running the simulation may or may not resemble the physics of the world that we observe.” At this point, we should ask: will the real physics please stand up! By “totality of physical existence,” he clearly means the *appearance* of what we call physical reality, which could be different in the simulation than it is for the simulators (who can create any simulated world they want, so long as it is not so inconsistent that it drives the simulated physicists mad!). By premise, the “physics” in the simulation (the world where presumptively we live) is a fiction and not the true physics of the real world. However, the inevitable corollary is that—if we cannot be sure that we do not live in a simulation—the same doubt must occur to our simulators concerning their world. The idea of a “fundamental level of reality” is meaningless if it is simulations all the way down!

Bostrom mines this dilemma for some moral precepts, such as a universal ethical imperative: “For example, if nobody can be sure that they are at the basement-level, then everybody would have to consider the possibility that their actions will be rewarded or punished... by their simulators. An afterlife would be a real possibility. Because of this fundamental uncertainty, even the basement civilization may have a reason to behave ethically.” It is precisely this uncertainty, of course, that is exploited by religion in our presumably real world. The quest of science is to find the “basement level” behind appearances. One of the fears that Darwin early inspired was that people might not behave ethically if they thought the world revealed by science is all that there is. If science is true, then physical reality is the basement level, and there is no afterlife and no punishing or rewarding gods. Then where would our moral direction come from? What incentive to behave well?

A strange further conclusion Bostrom draws is that *if* we indeed go on to create our own ancestor-simulations, this would be evidence that we ourselves live in a simulation. The reasoning is that doing so would be proof that we didn’t go extinct before reaching a post-human phase and would furthermore be evidence that we ourselves didn’t shy away from creating ancestor-simulations. (These are two factors that figure in his formula, implying that the percentage of people living in simulations must be large—and therefore probably includes us.) He also reasons that, if the cost of simulating a post-human society were prohibitively expensive for the hosts, we should expect the simulation we live in to be terminated soon—before we reach the post-human stage. (Then, if we do go extinct, will that be proof that we lived in a simulation all along?) This sort of reasoning in strange loops characterizes the Simulation Hypothesis, just as it does the self-verification of conspiracy theories. Indeed, it inheres in everything pertaining to the epistemic situation of the self-conscious mind!

Computational limits applied to gaming have given rise to the distinction between characters who are *players*, represented by an avatar within the game, and those characters that are non-players. They are both elements of the simulation, but the players are actual people within an existence outside the game. While Bostrom does not discuss gaming, his corresponding distinction is not between a real person and a simulation, but between the simulation of a real person (the ‘ancestor-simulation’ who is the supposed conscious subject experiencing the simulated world) and those other characters in the simulation who are mere props (the rest of humanity?). These serve as non-descript “others” for the conscious subject concerned; he calls “zombies” or “shadow people.” Obviously, the resources needed to support the simulated existence of a conscious subject have priority—if indeed there could be such a thing as simulated consciousness. But given that this

¹⁷ If we believe that God created the world, then we compete hopelessly against his presumably infinite resources!

hypothetical subject can choose to interact with these others in the game, in order to maintain the realism of the simulation and appear as “real” people or creatures, they too must potentially be instantly supported on demand. Then we have a recursive problem, since *those* subjects in turn may interact with props, and so on, creating an exponential demand on computational power. Bostrom belatedly notes that “It is not even obvious that it is possible for an entity to behave indistinguishably from a real human and yet lack conscious experience.” Indeed! The lesson in real life may be precisely that *interaction* with strangers leads us to take them seriously as subjects with an interior life of their own, to feel empathy for them rather than to treat them as mere objects or cyphers.

He concludes that the implications of living in a simulation “are not all that radical.” That seems a truism, given seamlessness of the illusion and the alleged probability that everyone else in the universe is similarly deluded. Yet, in the *Matrix* it makes all the difference, symbolized by the red pill and the blue pill. The premise of the film is to break out of the illusory simulation and find the real world, however dismal it may be, even as the simulators are determined to preserve the illusion. It would be entirely demoralizing to think that the so-called real world, found at last, turned out to be another simulation.

The last sentence of the paper shows that Bostrom’s ostensible purpose was not in fact to convince you that you might be living in a simulation or even to raise the question. Rather more modestly, it was to demonstrate—by contorted logic—that “unless we are now living in a simulation, our descendants will almost certainly never run an ancestor-simulation.” It *could* make a huge personal difference to know that one is living in a simulation created by superior beings, if that is even a meaningful concept. It is not really a personal concern what kind of simulations future people might create.

5. Discussion

The Simulation Argument presumes that we *could* live in a simulation and concludes that we probably do. The question posed and explored in Bostrom’s paper skirts the idea (which I take to be fact) that we *naturally* live in a simulation—or, more precisely, in a sort of virtual reality.¹⁸ The brain can be viewed as a control system; that constitutes a third-person perspective on the inputs and behavioral outputs of a biological organism. Yet, the brain can also be viewed as the natural agent (factory? computer? artist?) that produces our first-person consciousness as a sort of phenomenal show. I believe this show is functional; it serves a purpose which automated behavior cannot.¹⁹ It may seem to us that we are passive spectators to the panorama of consciousness—that is, the spectacle of the *world*—but truly we are agents, whose brains produce this show to help us survive and not especially to entertain us. In contrast, the virtual realities that we currently produce through computer simulation largely *are* entertainments. In any case, they *presume* a real embodied subject, using real senses to receive the input from virtual-reality devices. In future, it might be possible to “wire” a brain directly to a simulation program, bypassing the interface of the biological senses. This would be a physical realization of the brain in a vat. Even then, there still is a real embodied (not virtual) brain making sense (or not) of that input. And there is still a reality outside the simulation experienced by that brain—a reality that includes its own physical existence, as well as that of the vat, the computer running the simulation, and the real beings who built and run the computer. What we *mean* by simulation or virtual reality is a concept grounded in the real reality of our physical embodiment and actual technology. A fake reality presumes a real one. While it seems

¹⁸ A simulation attempts to model a presumed reality that is singular. A virtual reality is an original creation that may or not be based on the specific laws of the real world; there can be any number of arbitrary virtual worlds, so long as any one of them is sufficiently coherent to constitute a ‘world’ at all. Presumably, people live in slightly different virtual worlds because—while the objective world they have in common is singular—their brains vary slightly. But that, of course, is a realist presumption!

¹⁹ See my paper “Walking in the Shoes of the Brain” [<https://philpapers.org/rec/BRUWIT-8>]

a logical possibility, the notion of simulations all the way down, with no base reality, is a paranoid thought grounded in the inescapable paradox of being a self-conscious organism living in a hall of mirrors. While there is no way to *prove* that we are not living in an unnatural computer simulation, this is a different issue from living in the natural simulation provided by the brain.

Note also that a simulation is traditionally modelled on some reality, for some purpose. The idea is to make a “realistic” copy. This includes the simulations produced for virtual reality experiences, which are intended to be true enough to life to be convincing as entertainments. In contrast, the virtual reality produced by the brain is in no way a *copy*, but an original creation guided by sensory input from the (presumably) real external world. (Indeed, no copy of anything natural *can* be exhaustive; digital copies of digital files can be copied perfectly because the file is an artifact to begin with, not a natural entity.) We might think rather of reports from a news room: stories based on real events, with editorial slant. Realism in regard to the brain’s simulation has an entirely different meaning than *resemblance*. The virtual reality we naturally live in is realistic if it allows us to exist as products of natural selection.

An already developed natural brain could be “envatted” after the fact. But could it develop, as it were, *in silico*? A natural brain is embodied, which means it has a specific *relationship* to its environment programmed by natural selection. Can that relationship of embodiment by digitally simulated? If the experience of “realness” involves a functional relation of subject to object, deriving from an evolutionary contest, can it be effectively simulated by simulating the contest? If so, then it makes no difference whether one is living in a simulation; everything would unfold as it does in the real world. If *perfect* simulation is possible, then by definition there is no difference between simulation and reality, since the natural world could be perfectly replicated in a computer. But the question, then, is whether *perfect* simulation is possible. If “natural reality” *in principle* cannot be exhaustively mapped, then the answer is clearly no: perfect simulation is impossible for the same reasons that no final and exhaustive theory of everything is possible.²⁰

If it is intended to convincingly imitate natural reality, a simulation must involve a theory of physics, which would always be incomplete compared with natural reality. But if a simulation is intended merely to create an original fantasy, with its own “laws of nature” simply posited, a “glitch” would not be a deviation from “true” reality but no more than an internal inconsistency. It would either be a computer error or a result of the host programmer’s effort to correct an error.

More questions arise: what could it mean for a fictional subject to embrace a fictional world as *real*? And what would it matter for this fictional subject to conclude from detected glitches that its fictional world is a simulation and *not* real? To a real embodied subject, the difference between reality and fantasy can be a matter of life or death. What can it mean to a simulated subject with nothing (no body) at stake?

Absence of glitches is evidence, but not proof, of reality. The apparent self-consistency of the natural world could, after all, be due to the perfectness of an unnatural simulation. But there is certainly inconsistency among the perceptions of natural subjects. Is that a desirable feature to incorporate in a simulation for the sake of realism? It could be—if the goal is to emulate nominal real life in a way that would be convincing to artificial subjects modeled on human beings. But, again, why try to imitate a default reality if you can create any one you like?

On the other hand, there *are* discrepancies in our models of physical reality. We normally assume that this is because these models are incomplete, merely approximate, or wrong on some level. But, could they be mere glitches in a simulated world? For example, could the wave-particle duality be the sort of inconsistency that indicates life in a simulation? Perhaps “Bell’s no-go theorem, the EPR paradox, Bell’s inequalities...and quantum entanglement” can be explained by assuming we live in a simulation, where the simulators can do arbitrary things outside the classical laws of physics?²¹ But is such a thought an improvement over its religious version: that God can

²⁰ See my essay “Is Physical Reality Reducible to Thought?” [<https://stanceofunknowing.com/collected-writings/>]

²¹ See Campbell et al, op cit (my footnote 12). Like Barrow, their proposal for testing the Simulation Hypothesis by means of the two-slit experiment depends on so many implausible assumptions that it can hardly be deemed a test.

declare into existence what he likes, with no obligation to be consistent? Perhaps our simulators simply couldn't be bothered to make General Relativity compatible with quantum physics in their simulation?

A paper²² by John Barrow asks “what we might expect to see if we made scientific observations from within a simulated reality?” It is a question loaded with assumptions. First and foremost is the premise that “technical civilisations, only a little more advanced than ourselves, will have the capability to simulate universes in which self-conscious entities can emerge and communicate with one another.” Barrow’s question does not identify the “we” involved—whether it refers to real or simulated scientists. The Simulation Hypothesis trades on that very ambiguity. The notion of simulated *subjects* has already been deflated above. But granted that *we* (real biological subjects) have emerged as self-conscious entities, it may still seem a logical possibility that the world we experience could somehow (as in Descartes or *The Matrix*) be fake. Then the question remains: from the hypothetical simulated conscious epistemic subject’s point of view, is there a recognizable difference between a computer glitch in a simulation and a discrepancy between theory and observed reality? Ontologically, there is a clear distinction between the simulation and the reality simulated, since they have different histories and causes. The problem is the challenge for the subject to discriminate between them when the ontology itself is in question. But, circularly, this epistemic dilemma presumes one ontological reality or another, if it does not hopelessly muddle them together.²³ The observer confronting a simulated world is either real or simulated. If one remains *completely* skeptical (that is, without embracing this presumption) the problem cannot even be formulated.

In the case of science in the putatively real world, what Barrow charmingly calls the “flaws of nature” are gaps between physical reality and theory, which might be computer glitches. Unless the real world happens to *be* a finite structure whose complexity can eventually bottom out (in effect, a simulation), there will always be such gaps in our knowledge, simply because thought can never fully capture reality. In a simulated world, known to be generated by algorithms, there is no basis for this kind of gap or discrepancy.²⁴ For, a virtual world *can* be exhaustively described.²⁵ Any perceived “flaws” would indicate internal inconsistencies or random errors in the program, not the incomplete quest of real observers to discover the laws of their world. Were they feasible, *simulated observers* in a simulated world, who knew themselves to be such, would have no occasion to expect a difference between theory and reality. They would have no reason to conceive a difference between simulation and reality—aside from conclusions they might infer from their own ability to create simulations within their simulated world. Rather, they would be idealists who logically conclude that reality itself consists of no more than thought or imagination, so that there is no fundamental difference between reality and simulation.

5. Conclusion

It might have been more appropriate to call this essay a commentary than a refutation, since I haven’t strictly proven or disproven anything at all. In any case, credit is due to Nick Bostrom for

²² John D. Barrow “Living in a Simulated Universe” 2007 <https://www.simulation-argument.com > barrowsim.pdf>

²³ Indeed, any epistemology presumes an ontology—and vice-versa—since subject and object are inextricably entangled. But this particular dilemma presumes what it concludes by confounding distinct ontological realities.

²⁴ The laws of physics are sometimes likened to computer algorithms, and then falsely believed to generate physical reality. In truth, laws are no more than algorithmic compressions of data, after the fact, with no “governing power.”

²⁵ If, that is, it is not continually updated by unpredictable agents. Barrow points out the possibility that the hosts could correct internal errors on an ad hoc basis, keeping ahead of the simulated scientists’ ability to detect them: “This is a solution that will be very familiar to the owner of any home computer who receives regular updates in order to protect it against new forms of invasion or repair gaps that its original creators had not foreseen.” We might compare this to Newton’s idea that God must periodically intervene to reset the world machine. Barrow’s implied ontology includes real hackers, which reminds us again of Descartes’ paranoid vision.

articulating an issue that has been tacitly gnawing at the back of the human mind for centuries, and which erupted into the popular imagination with *The Matrix* in 1999, the very year that many believed, like millenarians of the past, would be the end of the world as we know it. Intentionally or not, his analysis reveals some of the treacherous convolutions involved in trying to understand our human epistemic situation from within itself, exacerbated in the digital age.

It can be a source of anxiety to wonder whether we live in a dream or simulation or whether there even is a real reality. But it can also be stimulating and fun. Above all, quite apart from the novelty of computer-generated virtual realities, the realization that we live in the *brain's* natural simulation is profound, with many epistemic and moral consequences. But that is another story.

On the other hand, if the question is unanswerable in any satisfying way, then perhaps it is a waste of time to ponder. I don't think so, because the consequences of *not* pondering it might cause even greater anxiety. If it serves us to ignore issues that arouse anxiety, the question is to what lengths we should go to avoid those issues and at what cost. Religion is one strategy to disarm the fundamental mystery of existence. Science may be too. Dogma of any sort provides immediate relief of anxiety and excuses us from the effort to make our own inquiries. On the other hand, it is dangerous when belief strays too far from reality. Ironically, the notion of a superior being (or beings) who create and maintain the world we experience, with absolute power over this world and its inhabitants, seems to be the very paradigm on which the Simulation Hypothesis is modeled. Religious or not, just getting on with life (whether real or simulated!) is what we mostly do, with the consequence that humanity has bumbled into an existential crisis that may be its doom. Ironically, then, avoiding Bostrom's question could inadvertently lead to failure to reach a post-human future. If that happens to be a desirable future—or at least a survivable one—then he has done us a service simply by reminding us of the question.