

# Chapter 2: Are We Living in the Matrix?

Sci Fi and Philosophy: Course Notes | Brendan Shea, PhD ([Brendan.Shea@rctc.edu](mailto:Brendan.Shea@rctc.edu))

In this chapter, we'll be discussing some philosophical issues that arise in the "Matrix" movie (and the larger "universe" of that movie). When the movie first came out (in 1999), the creators commissioned several famous philosophers, science fiction authors, artists, and others to give their own thoughts. Here, we'll look at some of the results.

## 1 CONTENTS

---

2	Movie Summary: The Matrix (1999).....	<b>Error! Bookmark not defined.</b>
3	Reading: Why Make a Matrix? And Why You Might Be In One (Nick Bostrom) .....	3
3.1	The Purpose of the Matrix.....	3
3.2	The Simulation Argument.....	5
3.3	Building a Matrix .....	5
3.4	Three Possibilities.....	6
3.5	Not the Old Brain-in-a-Vat Argument .....	6
3.6	How Could You Tell If You Are In A Matrix? .....	7
3.7	How To Live In A Matrix.....	8
3.8	Matrices Repeated and Stacked.....	9
4	Reading: Goliath (by Neil Gaiman) .....	10
4.1	Goliath.....	11
4.2	Thought Questions .....	19

## 2 INTRODUCTION TO THE MATRIX

---

The *Matrix* is a 1999 film directed by the Wachowskis. In this chapter, we'll be considering some philosophical themes that come in the movie. To start, however, it will help to have a brief overview of the plot, setting, and characters.

**Setting:** The movie is set in a dystopian future (in 2069) where humans are trapped in a simulated reality called the Matrix. The Matrix is created and maintained by artificial intelligences (AIs) who use humans as energy sources. However, to the characters in the movie, it "appears" to be 1999 (the year the movie is released).

**Characters:** The main characters in the movie are:

- **Neo** is the hero of the movie. He is a computer programmer who learns the nature of the Matrix, and uses his unique powers to help fight its creators.
- **Morpheus** is rebel leader who believes that Neo is the key to saving humanity
- **Trinity** is a skilled fighter, close friend of Morpheus, and love interest of Neo.
- **Cypher** is a member of Morpheus' team who betrays them in exchange for a comfortable life inside the Matrix
- **Agent Smith** is a program that is tasked with hunting down and eliminating the rebels
- **The Architect** is the AI creator of the Matrix who designed it to control humanity
- **The Oracle** is a program that helps guide Neo on his journey and helps him choose the right path

**Plot:** The movie is set in the year 2069 and revolves around a group of humans who are living in a simulated reality known as the Matrix. These humans are unaware of the true nature of their existence and are controlled by the machines that created the Matrix. One man, Neo, is contacted by a rebel group led by Morpheus and Trinity who tell him that the Matrix is not real and that he must help them overthrow the machine government. Neo is reluctant to believe this at first but eventually comes to accept it as the truth (and takes the famous “red pill”). He joins the rebels in their fight against the machines and eventually defeats them.

**Philosophical Themes.** Some important themes include the following:

- *The nature of reality*—The humans living in the Matrix are deceived about the “real nature” of the world around them. They *think* they live in houses, have families, work jobs, etc. However, they are actually just “brains in a vat.” This raises the question of how *we* (in the “real world”) can be so certain that we are NOT in this same position. That is: how do you know that you aren’t “in the Matrix” right now?
- *The nature of the mind*—In the Matrix, at least some of the AIs seem to have human-like abilities to reason and perceive. Moreover, they also seem to have *emotions*, with some (like Agent Smith) being angry and hateful, and others (The Oracle) being somewhat more friendly. This raises a few questions. First, are such machines possible and, if they are, what does this mean for our moral obligations toward them (for example, would it be OK for Neo and his friends to “kill all the machines” even if there were “innocent” machines)? Second, assuming such machines are possible (at least theoretically), how could we figure this out “from the outside”? Is there any test for determining whether a human/machine (remember, in the Matrix, you’re never sure which you are talking with) *really* has feelings, or is just pretending to have such feelings?
- *The nature of knowledge*—What does it mean to “know” something in the world of the Matrix? For example, suppose I am an ordinary human who believes that I “know” things like my own name, the location of the nearest grocery store, the fact that I have certain family and friends, etc. If I find out later that I am living in the Matrix, what does this mean for my beliefs? Do I not know *anything*? Or does much of my knowledge remain “intact”? (For example, I still “know” that I have a son—it’s just that he and I are both living in a world much different than the one we imagined!).
- *The nature of morality*—The Architect seems to believe (rightly or wrongly) that the Matrix he has created makes humans as “happy” as they could be, given the sorts of creatures they are. In particular, it makes them happier than they *would be*, if they knew the “truth about the world”. Suppose the Architect is right about this. Does this make designing the Matrix a “morally OK” thing to do? Why or why not? What about the rebel’s quest to destroy the Matrix, when many humans would prefer to continue living in it?
- *The meaning of life*—Unlike humans in the real world, the humans that live in the Matrix were “born” for a very specific purpose—to serve as energy sources for AIs. Moreover, as noted above, many of

them are perfectly happy living in this world. However, for characters like Neo, Trinity, and Morpheus, neither happiness nor “purpose” provides an answer to the question “What makes for a meaningful human life?” If the “meaning of life” is NOT (1) being happy or (2) doing whatever God/evolution/your parents “designed” you to do, what is it?

[Brendan: Have you seen the Matrix, or any of its sequels? Which scenes or events did you find especially memorable?]

### 3 READING: WHY MAKE A MATRIX? AND WHY YOU MIGHT BE IN ONE (NICK BOSTROM)<sup>1</sup>

---

**About the Author (Adapted from <https://nickbostrom.com/>).** Nick Bostrom is a Swedish-born philosopher with a background in theoretical physics, computational neuroscience, logic, and artificial intelligence, as well as philosophy. He is the most-cited professional philosopher in the world under the age of 50.

He is a Professor at Oxford University, where he heads the Future of Humanity Institute as its founding director. He is the author of some 200 publications, including *Anthropic Bias* (2002), *Global Catastrophic Risks* (2008), *Human Enhancement* (2009), and *Superintelligence: Paths, Dangers, Strategies* (2014), a New York Times bestseller which helped spark a global conversation about the future of AI. He has also published a series of influential papers, including ones that introduced the simulation argument (2003) and the concept of existential risk (2002).

Bostrom’s academic work has been translated into more than 30 languages. He is a repeat main TED speaker and has been interviewed more than 1,000 times by various media. He has been on Foreign Policy’s Top 100 Global Thinkers list twice and was included in Prospect’s World Thinkers list, the youngest person in the top 15. As a graduate student he dabbled in stand-up comedy on the London circuit, but he has since *The Purpose of the Matrix*

Why the Matrix? Why did the machines do it? (Human brains may be many things, but efficient batteries they are not.) How could they justify a world whose inhabitants are systematically deceived about their fundamental reality, ignorant about the reason why they exist, and subject to all the cruelty and suffering that we witness in the world around us? Children dying of AIDS; lovers separated by war and poverty; cancer patients tormented by unbearable pain; stroke victims deprived of their use of language and reason. One would think nobody but a sadist could have the imagination to think up these horrors, much less possess the desire to create a world that contains them in such abundance. But the machines did it, at least that’s how the story goes.

Although the world of the Matrix they created is far from perfect, it is arguably better than no world at all, the elimination of all human beings. Still, the machines could have created a world containing much more goodness, happiness, wisdom, personal growth, love and beauty, a world that was free of most of the natural and manmade evil that pervades our world. Indeed, as the story goes, they tried that, but supposedly it didn’t work.

***Agent Smith:*** *Did you know that the first Matrix was designed to be a perfect human world. Where none suffered. Where everyone would be happy. It was a disaster. No one would accept the program. Entire crops were lost. Some believed that we lacked the programming language to describe your perfect world. But I believe*

---

<sup>11</sup> From: *More Matrix and Philosophy: Revolutions and Reloaded Decoded*, ed. William Irwin (Open Court, 2005).

*that as a species, human beings define their reality through misery and suffering. The perfect world was a dream that your primitive cerebrum kept trying to wake up from. Which is why the Matrix was redesigned to this, the peak of your civilization*

The existence of unnecessary evil is one of the most powerful arguments against the belief that the world was created by an all-powerful, all-knowing, and perfectly good God. Theologians have spent centuries trying to answer it, and with very questionable success. But the problem of evil is only a problem if one assumes that the world was created by an omnipotent and perfectly good being. If one assumes instead that the creator was not perfectly good, and perhaps not even omnipotent, then it would be much easier to reconcile the view that our world was created with its seemingly obvious ethical shortcomings.

reconnected with the heavy gloom of his Swedish roots.

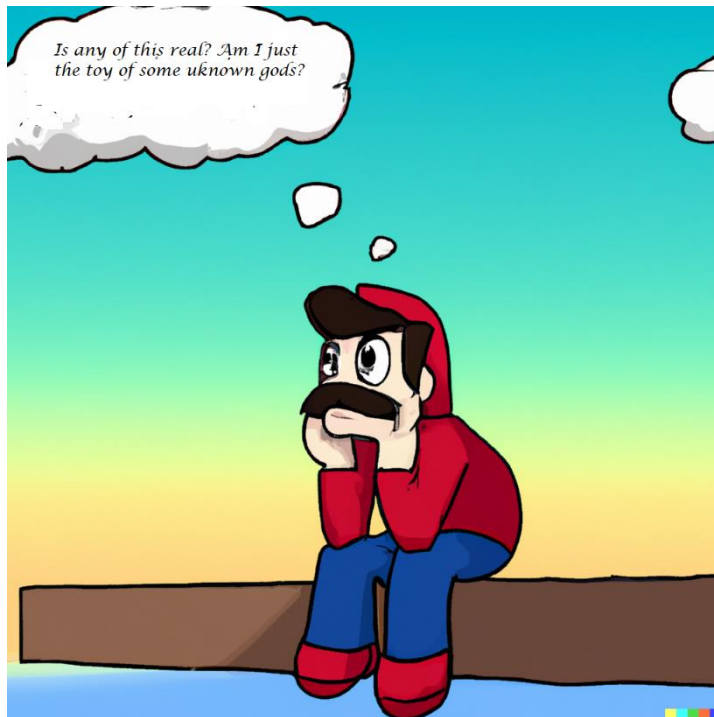


Figure 1 Super Mario and the meaning of life (Brendan Shea x Dall-E).

What about you? You're not all-powerful, all-knowing, and perfectly good. But what if you had the ability to create this kind of Matrix, would you do it? Even if *you* would not have chosen to create a world like this, there are many other people who do not share your scruples. If these people had the ability to create Matrices, some of their works might well look like the world in which we find ourselves.

Why might they choose to build a Matrix like our reality? One can think of many possible reasons, setting aside the daft idea of using human brains as batteries. But perhaps future

historians would create a Matrix that mimicked the history of their own species. They might do this to find out more about their past, or to explore counterfactual historical scenarios. In the world of the Architect(s), Napoleon may have succeeded in conquering Europe, and our world might be a Matrix created to research what would have happened if Napoleon had been defeated. Or perhaps there will be future artists who create Matrices as an art form much like we create movies and operas. Or perhaps the tourist industry will create simulations of interesting historical epochs so that their contemporaries can go on themed holidays to some bygone age by entering into the simulation and interacting with its inhabitants. The possible motives are myriad, and if future people are anything like present people, and if they have the technological might and the legal right to create Matrices, we would expect that many Matrices would be created, including ones that would look like the world that we are experiencing.

**[Brendan: Suppose that you were in charge of designing a “Matrix” in which billions of people would live, and that your only goal was to make these people’s lives as “good” as possible. To what extent would you make it similar to/different than the way the way our world is?]**

### 3.1 THE SIMULATION ARGUMENT

If each advanced civilization created many Matrices of their own history, then most people like us, who live in a technologically more primitive age, would live inside Matrices rather than outside them. If this were the case, where would you most likely be?

The so-called Simulation argument, which I introduced a few years ago, makes this line of reasoning more precise and takes it to its logical conclusion. The conclusion is that there are three basic possibilities at least one of which is true. The first possibility is that the human species will almost certainly go extinct before becoming technologically mature. The second possibility is that almost no technologically mature civilization is interested in building Matrices. The third possibility is that we are almost certainly living in a Matrix. Why? Because if the first two possibilities are not the case, then there are more *people* living in Matrices than in “real worlds.” As a person then the chances are that you are living in a Matrix rather than in a real world.

The Simulation argument does not tell us which of these three possibilities obtain, only that at least one of them does. The argument employs some math and probability theory, but the basic idea can be understood without recourse to technical apparatus.[\[1\]](#)

**[Brendan: To be clear: The simulation argument (which we’ll learn about in more detail) entails that YOU are likely to be living in a computer simulation right now. Before starting, how plausible do you think this is?]**

### 3.2 BUILDING A MATRIX

Creating comprehensive Matrices that are indistinguishable from non-simulated reality is, of course, far beyond our current technological capability. Even so, we can estimate the computational requirements for creating such virtual realities.

Rather than confining the construction project to creating a virtual reality simulation, we can consider a more ambitious project that also involves the creation of the inhabitants of the Matrix. Instead of having pink gooey pods with biological humans floating in them being fed sensory input from a simulated reality, it would be more efficient to replace the brains with simulations of brains. **Many philosophers and cognitive scientists believe that such brain-simulations would be conscious, provided the simulation was sufficiently detailed and accurate. [Brendan: Do you agree? Why or why not?]**

Estimates of the human brain’s computational power have been given and estimates of the computational power that would be available to a technologically mature civilization can also be made. While these estimates are very approximate, it turns out that even when allowing for a large margin of error, the computational resources of a mature civilization would suffice to create very many Matrices. Even a single planetary-sized computer, constructed with advanced molecular nanotechnology, could simulate the entire mental history of humankind by using less than one millionth of its computing power for one second; and this presupposes only already known computational mechanisms and engineering principles. A single civilization may eventually build millions of such computers. We can conclude that a technologically mature civilization would have enough computing power such that even if it devoted but a tiny fraction of it to creating Matrices, there would soon be many more simulated people than there were people living in the original history of that civilization.

These simulations would not have to be perfect. They would only have to be good enough to fool its inhabitants. It would not be necessary to simulate every object down to the subatomic level (something that would definitely be infeasible). If the book you are holding in your hands is a simulated book, the simulation would only need to include its visual appearance, its weight and texture, and a few other macroscopic properties, because you have no way of knowing what its individual atoms are doing at this moment. If you

were to study the book more carefully, for example by examining it under a powerful microscope, additional details of the simulation could be filled in as needed. Objects that nobody is perceiving could have an even more compressed representation. Such simplifications would dramatically reduce the computational requirements.

**[Brendan: Here, Bostrom argues that we will—in the not-so-distant future—have the computer hardware necessary for running simulations containing “virtual” human minds that could see/experience their world in much the same way we see/experience ours. What do you think?]**

### 3.3 THREE POSSIBILITIES

Given that the Architects of a technologically mature civilization could create a vast number of Matrices even by devoting just a small fraction of their resources to that end, an interesting implication follows. Consider the set of civilizations that are at similar level of technological development as our own current civilization. Suppose that some non-trivial fraction of these eventually go on to become technologically mature. Suppose, furthermore, that some non-trivial fraction of these devote a non-negligible proportion of their resources to building Matrices. Then most people like us live in Matrices rather than outside them. There are thus three basic possibilities: *either* almost every civilization like ours go extinct before reaching technological maturity; *or* almost every mature civilization lacks any interest in building Matrices; *or* almost all people with our kind of experiences live in Matrices.

Let us think a little about these three possibilities. If almost every civilization at our current stage goes extinct before becoming technologically mature, then our future looks relatively bleak. For if such a premature ending were the fate awaiting most civilizations, we would have to suspect that the same will hold for our civilization in particular. This is because we seem to lack any reason for thinking that our civilization will be luckier than most other civilizations at our stage.

The second possibility is less depressing. It might turn out that almost all technologically mature civilizations lose interest in building Matrices. Maybe the potential Architects of the future will not share any of the possible motives for building Matrices that we discussed above. Presumably, Architects would have used their advanced technology to improve their own capacities, so they may be superintelligent and have complete control over their own mental states. Rather than resorting to Matrix-building for recreation, they may obtain pleasure more efficiently by direct stimulation of their brains' pleasure centers. Their science may be so advanced that they have little to learn from running simulations of their historical past. Furthermore, they might develop ethical norms that prohibit the creation of Matrices. So we cannot infer from the fact that many *current* people would be tempted to construct Matrices that the same would hold for the super-advanced folks that would actually have the ability to act on this motive.

The third possibility is the most intriguing. If the vast majority of all people with other kind of experiences live in Matrices then *we* probably live in a Matrix. Unless we had some specific evidence to the contrary, we would therefore have to conclude that the world we see around us exists only by virtue of being simulated on a powerful computer built by some technologically highly advanced Architect.

**[Brendan: Which of the three possibilities for our future civilization do you think is most likely? Least likely? Why?]**

### 3.4 NOT THE OLD BRAIN-IN-A-VAT ARGUMENT

For hundreds of years, philosophers have pondered the question how we can know that the external world exists. Descartes (1596-1650) posed this question in his *Meditationes*, and considered the scenario where a hypothetical evil demon caused us to have erroneous beliefs about external objects. In more recent years,

Descartes' skeptical scenario has been given a more modern finish, and instead of a demon one is now asked to imagine a mad scientist who has extracted one's brain and who keeps it in a vat where the scientist is stimulating it with electrical signals replicating the sensory input that the brain would have had if it had interacted with a very different environment from that which is present in the real world. This is, of course, is the predicament explored in the Matrix movie. How can one possibly know that one is not such a brain in a vat, the philosophical skeptic challenges, given that all the appearances we experience could be the experiences of an envatted brain?

The argument outlined above provides a much stronger reason for taking seriously the possibility that we are living in a Matrix. The traditional skeptical argument offers no positive ground for thinking that we are living in a Matrix. At best, it shows that we cannot completely rule out that possibility, but we remain free to assign it a very small or negligible probability. If there are no mad scientists who experiment on conscious envatted human brains, then we are not envatted. Even if there were a few such brains-in-vats, they might be extremely rare compared to the brains-in-crania that interact with the external world in the normal way; and if so, then it may be highly unlikely that we would be among the envatted ones.

The Simulation argument, by contrast, adopts as its starting point that things are the way they seem to be and that science gives us reliable information about the world. Part of this information concerns the technological capabilities that an advanced civilization would be able to develop. Among these would be the capability to create Matrices. Crucially, it seems that they could easily create Matrices in astronomical numbers. From this we can then conclude that *either* technologically mature civilizations that are interested in creating Matrices are extremely rare compared to civilizations at our own current stage of development *or* almost all people like us live in Matrices. And from this, the division into three the three basic possibilities mentioned above follows.

The Simulation argument itself doesn't tell us which one of these three possibilities obtain. In fact, we do not currently have any strong evidence either for or against either of these three possibilities. We should therefore assign them all a significant probability. In particular, we should take seriously the possibility that we are living in a Matrix. We might still think that the probability is less than 50%. A degree of belief of something like 20% would seem quite reasonable given our current information.

**[Brendan: Bostrom's point here about probabilities is important! As it turns out, his Simulation argument (which concludes "you are living in a simulation") works so long as you agree there is *some* significant probability that technologically advanced civilizations would create Matrices.]**

### 3.5 HOW COULD YOU TELL IF YOU ARE IN A MATRIX?

Consider the predicament of Neo and his fellow rebels in the trilogy. They *know* there are many Matrices. They lead parts of their lives inside a Matrix. They know that most of their compatriots spend their whole lives in a Matrix. Given this, they should be extremely reluctant to think that they have escaped their Matrix. What appears to be an escape could easily just be simulated escape, so that they exit one level of the Matrix only to reemerge at another. The Wachowski brothers can of course stipulate that this is not the case and that the heroes really do get to experience "real" reality. But if Neo were rational, he would never be able to be at all confident that this is what happens.

If the Wachowski brothers had created a *real* Matrix (rather than just a movie *about* a Matrix), then, if they were rational, they would have to conclude that *they* themselves are almost certainly in a Matrix. For if we develop the capability to create our own Matrices, and if we decide to make use of this capability, we would obtain very strong evidence against the first two possibilities: that it is *not* the case that almost all civilizations at our current stage go extinct before reaching technological maturity and that it is *not* the case that almost

all mature civilizations lose interest in creating Matrices. This would leave us with only the third possibility—that we almost certainly inhabit a Matrix.

But what about the situation we actually find ourselves in? The Simulation argument aside, would it be possible to detect any direct signs of being in a Matrix? Is there a kind of “splinter in the mind” that would indicate that all is not right with reality? Certainly, if the Architects of a Matrix wished to reveal themselves, it would be easy enough for them to do so. For example, they could make a window pop up in our visual field with the text “YOU ARE LIVING IN A MATRIX. CLICK HERE FOR MORE INFORMATION”.

In the movie, the Oracle tells us that UFOs, Ghosts, and other strange sights are the manifestations of malfunctions in the Matrix that are being covered up.

***The Oracle:** Look, see those birds? At some point a program was written to govern them. A program was written to watch over the trees, and the wind, the sunrise, and sunset. There are programs running all over the place. The ones doing their job, doing what they were meant to do, are invisible. You'd never even know they were here. But the other ones, well, we hear about them all the time.*

***Neo:** I've never heard of them.*

***The Oracle:** Of course you have. Every time you've heard someone say they saw a ghost, or an angel. Every story you've ever heard about vampires, werewolves, or aliens is the system assimilating some program that's doing something they're not supposed to be doing.*

Deja vu is a sign of a glitch in the Matrix, which is re-running a sequence to cover something that has changed. Some people have written to me that they have found signs that we are in a Matrix. One person, for instance, told me that he could see flickering pixels when he looked in his bathroom mirror. Another person wrote that he could hear voices in his head. But even if we are in a Matrix, it is far more likely that such phenomena are the result of imperfections in the reporters rather than in the Matrix itself. There are many perfectly ordinary explanations for why some people should report having these kinds of experiences, including mental illness, over-excited imagination, gullibility, and so forth. Dysfunctional brains could be simulated just as easily as properly functioning ones, and including them in the simulation may indeed add to its verisimilitude.

Building any kind of Matrix at all that contains conscious simulated brains would be tremendously difficult. Any being capable of such a feat would almost certainly also be able to prevent any glitches in their Matrix from being noticed by its inhabitants. Even if some people did notice an anomaly, the Architect could backtrack the simulation a few seconds and rerun it in a way that avoided the anomaly entirely or else could simply edit out the memory of the anomaly from whoever had noticed something suspect.

**[Brendan: Come up with a possible “test” for determining whether you are currently living in a computer simulation? Now, consider how the “designers” of the computer simulation might manage to “trick” this test.]**

### 3.6 HOW TO LIVE IN A MATRIX

If we knew the Architects’ motives for designing Matrices then the hypothesis that we live in one might have major practical consequences. But in fact we know almost nothing about what these motives might be. Because of this ignorance, our best method for getting around in our Matrix (if that is where we are) is to study the patterns we find in the world we experience. We would run experiments, discover regularities, build models, and extrapolate from past events. In other words, we would apply the scientific method and common sense in the same way as if we knew that we were not in a Matrix. **To a first approximation, therefore, the**



**answer to how you should live if you are in a Matrix is that you should live the same way as if you are not in a Matrix.**

The Simulation argument does, however, have some more subtle practical ramifications, even if we set aside the other two possibilities to which it points (which do not entail that we are in a Matrix). Some scenarios that would otherwise seem to have been foreclosed by our current scientific understanding again become real possibilities if we inhabit a Matrix. For instance, while the physical world cannot suddenly pop out of existence, a simulated reality could do so at any time if the Architect decides to pull the plug. An afterlife would also be a real possibility. When a person dies in a simulation, he or she could be resurrected in another simulation, or the Architect could uplift the deceased into his own level of reality.

It is also conceivable that only some people are simulated in enough detail to be conscious while others may be simulated at a cruder level allowing them to appear and behave much like the real people but without having any subjective experience. The so-called *problem of other minds*—how we can know that other people are really conscious and are not just behaving as if they were—is another old chestnut of philosophy. There is, however, no consensus that such “zombie” people are possible even in principle. Some people have argued that it is necessarily true that anybody who acts sufficiently like a normal human being must also have conscious experience. (Whether this view would entail that your least favorite politicians cannot be zombies is a question on which more research is required.)

Another possibility is that the Architect might decide to reward or punish his simulated creatures, perhaps on the basis of moral criteria. If you might be in a Matrix, this consideration may give you a novel self-interested reason for behaving morally. The situation would be analogous to the case where God is watching and judging you except that the role of the final judge would not be a supernatural being but the physical person or persons who built the Matrix.

It would be misleading to say that if we are in a Matrix then we and the world around us do not really exist. It would be more accurate to say that the reality of these things is of a somewhat different nature than we thought before. Your nose would still be real; only, its reality would consist in being simulated on a powerful computer. The computer and the electrical activity of its circuitry would be physical phenomena in the more basic level of reality inhabited by the Architect of the Matrix.

**[Brendan: Suppose that you were fully convinced by Bostrom’s argument—you are currently living in a Matrix. How would this change the way you live your life, if at all?]**

### **3.7 MATRICES REPEATED AND STACKED**

When Neo stopped the Sentinels with his mind outside the Matrix at the end of *Reloaded* the speculation began. Was there a Matrix on top of the Matrix? As *Revolutions* revealed, there was not. But there could have been. A mature civilization would have enough computing power to run astronomically many Matrices. If we are in a Matrix, therefore, there are probably vast numbers of other Matrices, which differ from ours in some detail or in their overall design. These other Matrices may be run sequentially, as in the movie, or simultaneously by time-sharing the same processor or by using multiple computers. From the viewpoint of the simulated inhabitants, it makes little difference how the Matrices are implemented.

A Matrix may contain a civilization that matures and proceeds to build its own Matrices in the simulation. Reality could thus contain many levels, with computers being simulated inside computers which are themselves simulated, and so forth. How many layers of simulation there could be depends on the computing power available to the bottom-level Architect (who is not simulated). Since all the higher levels of simulation would ultimately be implemented on this Architect’s computer, he would have to shoulder the cost of all the

simulations and all the simulated people. If his computing power is limited, there may be only a small number of levels.



Figure 2 Figure 2 Suppose you found out you were "really" a robot. Would this change the way you feel about family/friends/etc. (Art by Brendan Shea x Dall-E).

As we noted above, all Architects would have strong reason to think that they themselves might be in a Matrix. (If the Architects at the basement level believed this they would be mistaken, but only because of bad epistemic luck, not because of any fault in their reasoning.) If we combine this insight with the speculation that moral considerations may play a part in determining the treatment some simulated people receive at the hands of their Architects, we are led to the peculiar thought that everybody—not just the simulated people—may have a self-interested reason for behaving morally. If behaving morally towards somebody includes judging and treating them according to moral criteria, this could further strengthen the reason that everybody have for behaving morally. The stronger that reason is, the more we would expect that people would be motivated by it. And the more people are likely to be motivated to treating their simulated creatures morally,

the stronger this reason would become. This reasoning can be iterated indefinitely in a truly **“virtuous circle,”** albeit a rather tenuous one as it relies not only on the possibility that we are in a simulation but also on tenuous speculations about the motives of the Architects.

At a minimum, the Simulation argument provides many exciting avenues for philosophical thinking. But if it is sound—and so far it has not been refuted—it could also provide various suggestions, however tentative and ambiguous, for how we should go about our lives and for what we should expect in the future. When we follow through the logical implications of what we think we know, we discover just how much we don’t yet know.

[Brendan: So, what do you think—does the Simulation Hypothesis provide us reason to behave ethically toward one another? Why or why not?]

## 4 READING: GOLIATH (BY NEIL GAIMAN)

**Brendan’s Note:** This is a short story (there is also a comic version) that was written as a tie-in to the Matrix movie. It was released on the Matrix movie website the same day the movie came out. Gaiman also provides a “local” link to the Matrix—he was living in Menominee, WI at the time he wrote this!]

**About the Author** (adapted from [https://www.neilgaiman.com/About\\_Neil/Biography](https://www.neilgaiman.com/About_Neil/Biography)) . Neil Gaiman was born in Hampshire, UK, and now lives in the United States near Minneapolis. As a child he discovered his love of books, reading, and stories, devouring the works of C.S. Lewis, J.R.R. Tolkien, James Branch Cabell, Edgar Allan Poe, Michael Moorcock, Ursula K. LeGuin, Gene Wolfe, and G.K. Chesterton. A self-described "feral child who was raised in libraries," Gaiman credits librarians with fostering a life-long love

of reading: "I wouldn't be who I am without libraries. I was the sort of kid who devoured books, and my happiest times as a boy were when I persuaded my parents to drop me off in the local library on their way to work, and I spent the day there. I discovered that librarians actually want to help you: they taught me about interlibrary loans."

Neil Gaiman is credited with being one of the creators of modern comics, as well as an author whose work crosses genres and reaches audiences of all ages. He is listed in the Dictionary of Literary Biography as one of the top ten living post-modern writers and is a prolific creator of works of prose, poetry, film, journalism, comics, song lyrics, and drama. His works include the comic book series *The Sandman* and novels *Stardust*, *American Gods*, *Coraline*, and *The Graveyard Book*. He has won numerous awards, including the Hugo, Nebula, and Bram Stoker awards, as well as the Newbery and Carnegie medals. He is the first author to win both the Newbery and the Carnegie medals for the same work, *The Graveyard Book* (2008).

## 4.1 GOLIATH

I suppose I could claim that I had always suspected that the world was a cheap and shoddy sham, a bad cover for something deeper and weirder and infinitely more strange, and that, in some way, I already knew the truth. But I think that's just how the world has always been. And even now that I know the truth—as you will, my love, if you're reading this—the world still seems cheap and shoddy. Different world, different shoddy, but that's how it feels.

They say, *Here's the truth*, and I say, *Is that all there is?* And they say, *Kind of. Pretty much. As far as we know.*

So. It was 1977, and the nearest I had come to computers was I'd recently bought a big, expensive calculator, and then I'd lost the manual that came with it, so I didn't know what it did anymore. I'd add, subtract, multiply, and divide, and was grateful I had no need to cos, sine, or find tangents or graph functions or whatever else the gizmo did, because, having recently been turned down by the RAF, I was working as a bookkeeper for a small discount carpet warehouse in Edgware, in north London, near the top of the Northern Line. I pretended that it didn't hurt whenever I'd see a plane overhead, that I didn't care that there was a world my size denied me. I just wrote down the numbers in a big double-entry book. I was sitting at the table at the back of the warehouse that served me as a desk when the world began to melt and drip away.

Honest. It was like the walls and the ceiling and the rolls of carpet and the News of the World topless calendar were all made of wax, and they started to ooze and run, to flow together and to drip. I could see the houses and the sky and the clouds and the road behind them, and then that dripped and flowed away, and behind it all was blackness.

I was standing in the puddle of the world, a weird, brightly colored thing that oozed and brimmed and didn't cover the tops of my brown leather shoes. (I have feet like shoeboxes. Boots have to be specially made for me. Costs me a fortune.) The puddle cast a weird light upward.

In fiction, I think I would have refused to believe it was happening, would've wondered if I'd been drugged or if I was dreaming. In reality, hell, I was there and it was real, so I stared up into the darkness, and then, when nothing more happened, I began to walk, splashing through the liquid world, calling out, seeing if anyone was about.

Something flickered in front of me.

"Hey fella," said a voice. The accent was American, although the intonation was odd.

"Hello," I said. The flickering continued for a few moments, and then resolved itself into a smartly dressed man in thick horn-rimmed spectacles.

“You’re a pretty big guy,” he said. “You know that?”

Of course I knew that. I was nineteen years old and even then I was close to seven feet tall. I have fingers like bananas. I scare children. I’m unlikely to see my fortieth birthday: people like me die young.

“What’s going on?” I asked. “Do you know?”

“Enemy missile took out a central processing unit,” he said. “Two hundred thousand people, hooked up in parallel, blown to dead meat. We’ve got a mirror going of course, and we’ll have it all up and running again in next to no time. You’re just free-floating here for a couple of nanoseconds, while we get London processing once more.”

**[Brendan: Note that in THIS version of the Matrix, human brains are used as computers, as opposed to energy sources. Which do you think is more plausible?]**

“Are you God?” I asked. Nothing he had said had made any sense to me.

“Yes. No. Not really,” he said. “Not as you mean it, anyway.”

**[Brendan: From a human perspective, what are some similarities between the machine creators of the Matrix and “God”?]**

And then the world lurched and I found myself coming to work again that morning, poured myself a cup of tea, had the longest, strangest bout of déjà vu I’ve ever had. Twenty minutes, where I knew everything that anyone was going to do or say. And then it went, and time passed properly once more, every second following every other second just like they’re meant to.

And the hours passed, and the days, and the years. I lost my job in the carpet company and got a new job bookkeeping for a company that sold business machines. I got married to a girl called Sandra I met at the swimming baths and we had a couple of kids, both normal sized, and I thought I had the sort of marriage that could survive anything, but I hadn’t, so she went away and she took the kiddies with her. I was in my late twenties, and it was 1986, and I got a job in a little shop on Tottenham Court Road selling computers, and I turned out to be good at it.

I liked computers. I liked the way they worked. It was an exciting time. I remember our first shipment of ATs, some of them with 40-megabyte hard drives.... Well, I was impressed easily back then.

I still lived in Edgware, commuted to work on the Northern Line. I was on the tube one evening, going home—we’d just gone through Euston and half the passengers had got off—and I was looking at the other people in the carriage over the top of the Evening Standard and wondering who they were, who they really were, inside: the thin, black girl writing earnestly in her notebook, the little old lady with the green velvet hat on, the girl with the dog, the bearded man with the turban.

And the tube stopped in the tunnel.

That was what I thought happened, anyway: I thought the tube had stopped. Everything went very quiet.

And then we went through Euston, and half the passengers got off. And then we went through Euston, and half the passengers got off. And I was looking at the other passengers and wondering who they really were inside when the train stopped in the tunnel, and everything went very quiet.

And then everything lurched so hard I thought we’d been hit by another train.

And then we went through Euston, and half the passengers got off, and then the train stopped in the tunnel, and then everything went—

*(Normal service will be resumed as possible, whispered a voice in the back of my head.)*

And this time as the train slowed and began to approach Euston I wondered if I was going crazy: I felt like I was jerking back and forth on a video loop. I knew it was happening, but there was nothing I could do to change anything, nothing I could do to break out of it.

The black girl sitting next to me passed me a note. ARE WE DEAD? it said.

I shrugged. I didn't know. It seemed as good an explanation as any. Slowly, everything faded to white. There was no ground beneath my feet, nothing above me, no sense of distance, no sense of time. I was in a white place. And I was not alone.

The man wore thick horn-rimmed spectacles, and a suit that looked like it might have been an Armani. "You again?" he said. "The big guy. I just spoke to you."

"I don't think so," I said.

"Half an hour ago. When the missiles hit."

"Back In the carpet factory? That was years ago. Half a lifetime."

"About thirty-seven minutes back. We've been running in an accelerated mode since then, trying to patch and cover, while we've been processing potential solutions."

"Who sent the missiles?" I asked. "The U.S.S.R.? The Iranians?"

"Aliens," he said. "You're kidding?"

"Not as far as we can tell. We've been sending out seed probes for a couple of hundred years now. Looks like something has followed one back. We learned about it when the first missiles landed. It's taken us a good twenty minutes to get a retaliatory plan up and running. That's why we've been processing in overdrive. Did it seem like the last decade went pretty fast?"

"Yeah. I suppose."

"That's why. We ran it through pretty fast, trying to maintain a common reality while coprocessing."

**[Brendan: So, "real" time (outside the Matrix) passes differently than "Matrix" time. From the point of view of humans, which matters? The "objective" passage of time? Or the "experienced" passage of time?]**

"So what are you going to do?"

"We're going to counterattack. We're going to take them out. I'm afraid it will take a while: we don't have the machinery yet. We have to build it."

The white was fading now, fading into dark pinks and dull reds. I opened my eyes. For the first time. I choked on it. It was too much to take in.

So. Sharp the world and tangled-tubed and strange and dark and somewhere beyond belief. It made no sense. Nothing made sense. It was real, and it was a nightmare. It lasted for thirty seconds, and each cold second felt like a tiny forever.

And then we went through Euston, and half the passengers got off....

I started talking to the black girl with the notebook. Her name was Susan. Several weeks later she moved in with me. Time rumbled and rolled. I suppose I was becoming sensitive to it. Maybe I knew what I was looking for—knew there was something to look for, even if I didn't know what it was.

I made the mistake of telling Susan some of what I believed one night—about how none of this was real. About how we were really just hanging there, plugged and wired, central processing units or just cheap memory chips for some computer the size of the world, being fed a consensual hallucination to keep us happy, to allow us to communicate and dream using the tiny fraction of our brains that weren't being used by them—whoever they were—to crunch numbers and store information. “We're memory,” I told her. “That's what we are. Memory.”

“You don't really believe this stuff,” she told me, and her voice was trembling. “It's a story.”

When we made love, she always wanted me to be rough with her, but I never dared. I didn't know my own strength, and I'm so clumsy. I didn't want to hurt her. I never wanted to hurt her, so I stopped telling her my ideas, tried to kiss it better, to pretend it had all been a joke, just not the funny kind....

It didn't matter. She moved out the following weekend.

I missed her, deeply, painfully. But life goes on.

The moments of déjà vu were coming more frequently now. Moments would stutter and hiccup and falter and repeat. Sometimes whole mornings would repeat. Once I lost a day. Time seemed to be breaking down entirely.

And then I woke up one morning and it was 1975 again, and I was sixteen, and after a day of hell at school I was walking out of school, into the RAF recruiting office next to the kebab house in Chapel Road.

**[Brendan: Why do you think this has happened to the main character?]**

“You're a big lad,” said the recruiting officer. I thought he was American at first, but he said he was Canadian. He wore big horn-rimmed glasses.

“Yes,” I said.

“And you want to fly?”

“More than anything.” It seemed like I half-remembered a world in which I'd forgotten that I wanted to fly planes, which seemed as strange to me as forgetting my own name.

“Well,” said the horn-rimmed man, “we're going to have to bend a few rules. But we'll have you up in the air in no time.” And he meant it, too.

The next few years passed really fast. It seemed like I spent all of them in planes of different kinds, cramped into tiny cockpits, in seats I barely fitted, flicking switches too small for my fingers. I got Secret clearance, then I got Noble clearance, which leaves Secret clearance in the shade, and then I got Graceful clearance, which the Prime Minister himself doesn't have, by which time I was piloting flying saucers and other craft that moved with no visible means of support.

I started dating a girl called Sandra, and then we got married, because if we married we got to move into married quarters, which was a nice little semi-detached house near Dartmoor. We never had any children: I

had been warned that it was possible I might have been exposed to enough radiation to fry my gonads, and it seemed sensible not to try for kids, under the circumstances: didn't want to breed monsters.

It was 1985 when the man with horn-rimmed spectacles walked into my house.

My wife was at her mother's that week. Things had got a bit tense, and she'd moved out to buy herself some "breathing room." She said I was getting on her nerves. But if I was getting on anyone's nerves, I think they must have been my own. It seemed like I knew what was going to happen all the time. Not just me: it seemed like everyone knew what was going to happen. Like we were sleepwalking through our lives for the tenth or the twentieth or the hundredth time.

I wanted to tell Sandra, but somehow I knew better, knew I'd lose her if I opened my mouth. Still, I seemed to be losing her anyway. So I was sitting in the lounge watching *The Tube* on Channel Four and drinking a mug of tea, and feeling sorry for myself.

The man with the horn-rimmed specs walked into my house like he owned the place. He checked his watch.

"Right," he said. "Time to go. You'll be piloting something pretty close to a PL-47."

Even people with Graceful clearance weren't meant to know about PL-47s. I'd flown a prototype a dozen times. Looked like a teacup, flew like something from *Star Wars*.

"Shouldn't I leave a note for Sandra?" I asked.

"No," he said, flatly. "Now, sit down on the floor and breathe deeply and regularly. In, out, in, out."

It never occurred to me to argue with him, or to disobey. I sat down on the floor, and I began to breathe, slowly, in and out and out and in and...

In.

Out.

In. A wrenching. The worst pain I've ever felt. I was choking.

In.

Out.

I was screaming, but I could hear my voice and I wasn't screaming. All I could hear was a low bubbling moan.

In.

Out.

It was like being born. It wasn't comfortable, or pleasant. It was the breathing carried me through it, through all the pain and the darkness and the bubbling in my lungs. I opened my eyes.

**[Brendan: How does this description of being "red-pilled" (waking up from the Matrix) compare to that shown in the movie, if you've watched it?]**

I was lying on a metal disk about eight feet across. I was naked, wet, and surrounded by a sprawl of cables. They were retracting, moving away from me, like scared worms or nervous brightly colored snakes.

I looked down at my body. No body hair, no scars, no wrinkles. I wondered how old I was, in real terms. Eighteen? Twenty? I couldn't tell.

There was a glass screen set into the floor of the metal disk. It flickered and came to life. I was staring at the man in the horn-rimmed spectacles.

"Do you remember?" he asked. "You should be able to access most of your memory for the moment."

"I think so," I told him.

"You'll be in a PL-47," he said. "We've just finished building it. Pretty much had to go back to first principles, come forward. Modify some factories to construct it. We'll have another batch of them finished by tomorrow. Right now we've only got one."

"So if this doesn't work, you've got replacements for me."

"If we survive that long," he said. "Another missile bombardment started about fifteen minutes ago. Took out most of Australia. We project that it's still a prelude to the real bombing."

"What are they dropping? Nuclear weapons?"

"Rocks."

"Rocks?"

"Uh-huh. Rocks. Asteroids. Big ones. We think that tomorrow, unless we surrender, they may drop the moon on us."

"You're joking."

"Wish I was." The screen went dull.

The metal disk I was riding had been navigating its way through a tangle of cables and a world of sleeping naked people. It had slipped over sharp microchip towers and softly glowing silicone spires.

The PL-47 was waiting for me at the top of a metal mountain. Tiny metal crabs scuttled across it, polishing and checking every last rivet and stud.

I walked inside on tree trunk legs that still trembled and shook from lack of use. I sat down in the pilot's chair and was thrilled to realize that it had been built for me. It fitted. I strapped myself down. My hands began to go through warm-up sequence. Cables crept over my arms. I felt something plugging into the base of my spine, something else moving in and connecting at the top of my neck.

My perception of the ship expanded radically. I had it in 360 degrees, above, below. I was the ship, while at the same time, I was sitting in the cabin, activating the launch codes.

"Good luck," said the horn-rimmed man on a tiny screen to my left.

"Thank you. Can I ask one last question?"

"I don't see why not."

"Why me?"

"Well," he said, "the short answer is that you were designed to do this. We've improved a little on the basic human design in your case. You're bigger. You're much faster. You have improved processing speeds and reaction times."





four inseparable organisms that dream they are one.

*Figure 3 Space octopus attacks (Brendan Shea x Dall-E).*

“I’m not faster. I’m big, but I’m clumsy.”

“Not in real life,” he said.

“That’s just in the world.”

And I took off.

I never saw the aliens, if there were any aliens, but I saw their ship. It looked like fungus or seaweed: the whole thing was organic, an enormous glimmering thing, orbiting the moon. It looked like something you’d see growing on a rotting log, half-submerged under the sea. It was the size of Tasmania.

Two-hundred mile-long sticky tendrils were dragging asteroids of various sizes behind them. It reminded me a little of the trailing tendrils of a Portuguese Man O’ War, that strange compound sea creature:

They started throwing rocks at me as I got a couple of hundred thousand miles away.

My fingers were activating the missile bay, aiming at a floating nucleus, while I wondered what I was doing. I wasn’t saving the world I knew. That world was imaginary: a sequence of ones and zeroes. If I was saving anything, I was saving a nightmare....

But if the nightmare died, the dream was dead, too.

**[Brendan: If this was you, would you agree to fight the aliens on behalf on the machines? Why or why not?]**

There was a girl named Susan. I remembered her from a ghost life long gone. I wondered if she was still alive. (Had it been a couple of hours ago? Or a couple of lifetimes?) I supposed she was dangling hairless from cables somewhere, with no memory of a miserable, paranoid giant.

I was so close I could see the ripples of the creature’s skin. The rocks were getting smaller and more accurate. I dodged and wove and skimmed to avoid them. Part of me was just admiring the economy of the thing: no expensive explosives to build and buy, no lasers, no nukes. Just good old kinetic energy: big rocks.

If one of those things had hit the ship I would have been dead. Simple as that.

The only way to avoid them was to outrun them. So I kept running.

The nucleus was staring at me. It was an eye of some kind. I was certain of it.

I was less than a hundred yards away from the nucleus when I let the payload go. Then I ran. I wasn’t quite out of range when the thing imploded. It was like fireworks—beautiful in a ghastly sort of way. And then there was nothing but a faint trace of glitter and dust....

“I did it!” I screamed. “I did it! I fucking well did it!”

The screen flickered. Horn-rimmed spectacles were staring at me. There was no real face behind them anymore. Just a loose approximation of concern and interest, like a blurred cartoon. “You did it,” he agreed.

“Now, where do I bring this thing down?” I asked.

There was a hesitation, then, “You don’t. We didn’t design it to return. It was a redundancy we had no need for. Too costly, in terms of resources.”

“So what do I do? I just saved the Earth. And now I suffocate out here?”

He nodded. “That’s pretty much it. Yes.”

The lights began to dim. One by one, the controls were going out. I lost my 360-degree perception of the ship. It was just me, strapped to a chair in the middle of nowhere, inside a flying teacup.

“How long do I have?”

“We’re closing down all your systems, but you’ve got a couple of hours, at least. We’re not going to evacuate the remaining air. That would be inhuman.”

“You know, in the world I came from, they would have given me a medal.”

“Obviously, we’re grateful.”

“So you can’t come up with any more tangible way to express your gratitude?”

“Not really. You’re a disposable part. A unit. We can’t mourn you any more than a wasps’ nest mourns the death of a single wasp. It’s not sensible and it’s not viable to bring you back.”

“And you don’t want this kind of firepower coming back toward the Earth, where it could potentially be used against you?”

“As you say.”

And then the screen went dark, with not so much as a good-bye. *Do not adjust your set*, I thought. *Reality is at fault.*

You become very aware of your breathing, when you only have a couple of hours of air remaining. In. Hold. Out. Hold. In. Hold. Out. Hold....

I sat there strapped to my seat in the half-dark, and I waited, and I thought. Then I said, “Hello? Is anybody there?”

A beat. The screen flickered with patterns. “Yes?”

“I have a request. Listen. You—you people, machines, whatever you are—you owe me one. Right? I mean I saved all your lives.”

“Continue.”

“I’ve got a couple of hours left. Yes?”

“About fifty-seven minutes.”

“Can you plug me back into the...the real world. The other world. The one I came from?”

“Mm? I don’t know. I’ll see.” Dark screen once more.

I sat and breathed, in and out, in and out, while I waited. I felt very peaceful. If it wasn’t for having less than an hour to live, I’d have felt just great.

The screen glowed. There was no picture, no pattern, no nothing. Just a gentle glow. And a voice, half in my head, half out of it, said, “You got a deal.”

There was a sharp pain at the base of my skull. Then blackness, for several minutes.

Then this.

That was fifteen years ago: 1984. I went back into computers. I own my computer store on the Tottenham Court Road. And now, as we head toward the new millennium, I’m writing this down. This time around, I married Susan. It took me a couple of months to find her. We have a son.

I’m nearly forty. People of my kind don’t live much longer than that, on the whole. Our hearts stop. When you read this, I’ll be dead. You’ll know that I’m dead. You’ll have seen a coffin big enough for two men dropped into a hole.

But know this, Susan, my sweet: my true coffin is orbiting the moon. It looks like a flying teacup. They gave me back the world, and you, for a little while. Last time I told you, or someone like you, the truth, or what I knew of it, you walked out on me. And maybe that wasn’t you, and I wasn’t me, but I don’t dare risk it again. So I’m going to write this down, and you’ll be given it with the rest of my papers when I’m gone. Good-bye.

They may be heartless, unfeeling, computerized bastards, leeching off the minds of what’s left of humanity. But I can’t help feeling grateful to them.

I’ll die soon. But the last twenty minutes have been the best years of my life.

## 4.2 THOUGHT QUESTIONS

1. How does the depiction of the Matrix compare to that in the movie?
2. The main character “knows” that he is living in the Matrix, and yet he clearly believes that his wife and children are “real” people worthy of care and affection. Why is this?
3. From the point of view of the main character, which world (“inside” the Matrix or “outside” the Matrix) is the “real” one?
4. If you were in the position of the main character, would you agree to fight the aliens? Would you ask to be plugged back into the Matrix? Why?
5. Suppose that you were given a choice of two lives: (1) a “normal” life in the Matrix where you never realize what’s happening, but get to live until 85, see your children grow up, etc. or (2) the life of the main character where you realize the “truth”, and end up dying at 40 (in Matrix time). Which would you choose? Why?

---

[1] For the full story, see (2003) “Are You Living In A Computer Simulation?” *Philosophical Quarterly*. Vol. 53, No. 211, pp. 243-255. This and other related papers are available at [www.simulation-argument.com](http://www.simulation-argument.com).