Reality+

by David Chalmers

excerpted from *Reality+ Virtual Worlds and the Problems of Philosophy* (2022)

Introduction
Adventures in Technophilosophy

When I was ten years old, I discovered computers. My first machine was a PDP-10 mainframe system at the medical center where my father worked. I taught myself to write simple programs in the BASIC computer language. Like any ten-year-old, I was especially pleased to discover games on the computer. One game was simply labeled "ADVENT." I opened it and saw:

> *You are standing at the end of a road before a small brick building.*
> *Around you is a forest.*
> *A small stream flows out of the building and down a gully.*

I figured out that I could move around with commands like "go north" and "go south." I entered the building and got food, water, keys, a lamp. I wandered outside and descended through a grate into a system of underground caves. Soon I was battling snakes, gathering treasures, and throwing axes at pesky attackers. The game used text only, no graphics, but it was easy to imagine the cave system stretching out below ground. I played for months, roaming farther and deeper, gradually mapping out the world.

It was 1976. The game was *Colossal Cave Adventure*. It was my first virtual world.

In the years that followed, I discovered video games. I started with *Pong* and *Breakout*. When *Space Invaders* came to our local shopping mall, it became an obsession for my brothers and me. Eventually I got an Apple II computer, and we could play *Asteroids* and *Pac-Man* endlessly at home.

Over the years, virtual worlds have become richer. In the 1990s, games such as *Doom* and *Quake* pioneered the use of a first-person perspective. In the 2000s, people began spending vast amounts of time in multiplayer virtual worlds like *Second Life* and *World of Warcraft*. In the 2010s, there arrived the first rumblings of consumer-level virtual reality headsets, like the Oculus Rift. That decade also saw the first widespread use of augmented reality environments, which populate the physical world with virtual objects in games like *Pokémon Go*.

These days, I have numerous virtual reality systems in my study, including an Oculus Quest 2 and an HTC Vive. I put on a headset, open an application, and suddenly I'm in a virtual world. The physical world has disappeared entirely, replaced by a computer-generated environment. Virtual objects surround me, and I can move among them and manipulate them.

Like ordinary video games from *Pong* to *Fortnite*, virtual reality (or VR) involves a virtual world: an interactive, computer-generated space. What's distinctive about VR is that its virtual worlds are *immersive.* Instead of showing you a two-dimensional screen, VR immerses you in a three-dimensional world you can see and hear as if you existed within it. Virtual reality involves an immersive, interactive, computer-generated space.

I've had all sorts of interesting experiences in VR. I've assumed a female body. I've fought off assassins. I've flown like a bird. I've traveled to Mars. I've looked at a human brain from the inside, with neurons all around me. I've stood on a plank stretched over a canyon—terrified, though I knew perfectly well that if I were to step off, I'd step onto a nonvirtual floor just below the plank.

Like many other people, during the recent pandemic I've spent a great deal of time talking to friends, family, and colleagues using Zoom and other videoconferencing software. Zoom is convenient, but it has many limitations. Eye contact is difficult. Group interactions are choppy rather than cohesive. There is no sense that we are inhabiting a common space. One underlying issue is that videoconferencing is not virtual reality. It is interactive but not immersive, and there is no common virtual world.

During the pandemic, I've also met up once a week with a merry band of fellow philosophers in VR. We've tried many different platforms and activities—flying with angel wings in *Altspace*, slicing cubes to a rhythm in *Beat Saber*, talking philosophy on the balcony in *Bigscreen*, playing paintball in *Rec Room*, giving lectures in *Spatial*, trying out colorful avatars in *VRChat.* VR technology is still far from perfect, but we've had the sense of inhabiting a common world. When five of us were standing around after a short presentation, someone said, "This is just like coffee break at a philosophy conference." When the next pandemic arrives in a decade or two, it's likely that many people will hang out in immersive virtual worlds designed for social interaction.

Augmented reality (or AR) systems are also progressing fast. These systems offer a world that is partly virtual and partly physical. The ordinary physical world is augmented by virtual objects. I don't yet have my own augmented reality glasses, but companies like Apple, Facebook, and Google are said to be working on them. Augmented reality systems have the potential to replace screen-based computing entirely, or at least replace physical screens with virtual screens. Interacting with virtual objects may become part of everyday life.

Today's VR and AR systems are primitive. The headsets and glasses are bulky. The visual resolution for virtual objects is grainy. Virtual environments offer immersive vision and sound, but you can't touch a virtual surface, smell a virtual flower, or taste a virtual glass of wine when you drink it.

These temporary limitations will pass. The physics engines that underpin VR are improving. In years to come, the headsets will get smaller, and we will transition to glasses, contact lenses, and eventually retinal or brain implants. The resolution will get better, until a virtual world looks exactly like a nonvirtual world. We will figure out how to handle touch, smell, and taste. We may spend much of our lives in these environments, whether for work, socializing, or entertainment.

My guess is that within a century we will have virtual realities that are indistinguishable from the nonvirtual world. Perhaps we'll plug into machines through a brain-computer interface, bypassing our eyes and ears and other sense organs. The machines will contain an extremely detailed simulation of a physical reality, simulating laws of physics to track how every object within that reality behaves.

Sometimes VR will place us in other versions of ordinary physical reality. Sometimes it will immerse us in worlds entirely new. People will enter some worlds temporarily for work or for pleasure. Perhaps Apple will have its own workplace world, with special protections so that no one can leak its latest Reality system under development. NASA will set up a world with spaceships in which people can explore the galaxy at faster-than-light speed. Other worlds will be worlds in which people can live indefinitely. Virtual real estate developers will compete to offer worlds with perfect weather near the beach, or with glorious apartments in a vibrant city, depending on what customers want.

Perhaps, as in the novel and movie *Ready Player One*, our planet will be crowded and degraded, and virtual worlds will provide us with new landscapes and new possibilities. In centuries past, families often faced a decision: "Should we emigrate to a new country to start a new life?" In centuries to come, we may face an equivalent decision: "Should we move our lives to a virtual world?" As with emigration, the reasonable answer may often be yes.

Once simulation technology is good enough, these simulated environments may even be occupied by simulated people, with simulated brains and bodies, who will undergo the whole process of birth, development, aging, and death. Like the nonplayer characters that one encounters in many video games, simulated people will be creatures of the simulation. Some worlds will be simulations set up for research or to make predictions about the future. For instance, a dating app (as seen on the TV series *Black Mirror*) could simulate many futures for a couple in order to see whether they are compatible. A historian might study what would have happened if Hitler had chosen not to start a war with the Soviet Union. Scientists might simulate whole universes from the Big Bang onward, with small variations to study the range of outcomes: How often does life develop? How often is there intelligence? How often is there a galactic civilization?

One can imagine that a few curious 23rd-century simulators might focus on the early 21st century. Let's suppose the simulators live in a world in which Hillary Clinton defeated Jeb Bush in the US presidential election of 2016. They might ask: How would history have been different if Clinton had lost? Varying a few parameters, the simulators might go so far as to simulate a world where the 2016 victor was Donald Trump. They might even simulate Brexit and a pandemic.

Simulators interested in the history of simulation might also be interested in the 21st century as a period when simulation technology was coming into its own. Perhaps they might occasionally simulate people who are writing books about possible future simulations, or people who are reading them! Narcissistic simulators might nudge the parameters so that some simulated 21st-century philosophers speculate wildly about simulations built in the 23rd century. They might be especially interested in simulating the reactions of 21st-century readers reading thoughts about 23rd-century simulators, as you are right now.

Someone in such a virtual world would believe themselves to be living in an ordinary world in the early 21st century—a world in which Trump was elected president, the UK left the European Union, and there was a pandemic. Those events may have been surprising at the time, but humans have a remarkable capacity to adjust, and after a while these things become normal. Although simulators may have nudged them into reading a book on virtual worlds, it will seem to them as if they are reading the book out of their own free choice. The book they're reading now is perhaps a little unsubtle in trying to convey the message that they may be in a virtual world, but they will take this in stride and start thinking about the idea all the same.

At this point, we can ask, "How do you know you're not in a computer simulation right now?"

This idea is often known as the *simulation hypothesis.* It is famously depicted in the *Matrix* movies, in which what seems an ordinary physical world turns out to be the result of connecting human brains to a giant bank of computers. Inhabitants of the Matrix experience their world very much as we do, but the Matrix is a virtual world.

Could you be in a virtual world right now? Stop and think about this question for a moment. When you do, you're doing philosophy.

Philosophy translates as *love of wisdom*, but I like to think of it as *the foundations of everything*. Philosophers are like the little kid who keeps asking, *Why?* or *What is that?* or *How do you know?* or *What does that mean?* or *Why should I do that?* Ask those questions a few times in a row and you rapidly reach the foundations. You're examining the assumptions that underlie things we take for granted.

I was that kid. It took me a while to realize that what I was interested in was philosophy. I started off studying mathematics, physics, and computer science. These take you a fair distance into the foundations of everything, but I wanted to go deeper. I turned to studying philosophy, along with cognitive science to keep an anchor in the solid ground of science while I explored the foundations underneath.

I was first drawn to address questions about the mind, like *What is consciousness?* I've spent much of my career focusing on those questions. But questions about the world, like *What is reality?*, are just as central to philosophy. Perhaps most central of all are questions about the relation between mind and world, such as *How can we know about reality?*

This last question was at the heart of the challenge posed by René Descartes in his *Meditations on First Philosophy* (1641), which set the agenda for centuries of Western philosophy to come. Descartes posed what I'll call the problem of the external world: How do you know anything at all about the reality outside you?

Descartes approached the problem by asking: How do you know that your perception of the world is not an illusion? How do you know that you are not dreaming right now? How do you know you're not being deceived by an evil demon into thinking all this is real, when it's not? These days, he might approach the problem by asking the question I just asked you: How do you know you're not in a virtual world?

For a long time I thought I didn't have much to say about Descartes's problem of the external world. Thinking about virtual reality gave me a new perspective. It was reflecting on the simulation hypothesis that led me to realize that I had underestimated virtual worlds. In their own way, so had Descartes and many others. I concluded that if we think more clearly about virtual worlds, this might lead us to the beginnings of a solution to Descartes's problem.

✦

This book is a project in what I call *technophilosophy*. Technophilosophy is a combination of (1) asking philosophical questions about technology and (2) using technology to help answer traditional philosophical questions. The name is inspired by what the Canadian-American philosopher Patricia Churchland called *neurophilosophy* in her landmark 1987 book of the same title. Neurophilosophy combines asking philosophical questions about neuroscience with using neuroscience to help answer traditional questions in philosophy. Technophilosophy does the same with technology.

There's a thriving area, often called the philosophy of technology, that carries out the first project—asking philosophical questions about technology. What's especially distinctive about technophilosophy is the second project—using technology to answer traditional philosophical questions. The key to technophilosophy is a two-way interaction between philosophy and technology. Philosophy helps to shed light on (mostly new) questions about technology. Technology helps to shed light on (mostly old) questions about philosophy. I wrote this book in order to shed light on both sorts of question at once.

✦

First, I want to use technology to address some of the oldest questions in philosophy, especially the problem of the external world. At a minimum, virtual reality technology helps *illustrate* Descartes's problem—that is, how can we know anything about the reality around us? How do we know that reality is not an illusion? In chapters 2 and 3, I lay out these problems by introducing the simulation hypothesis and asking, "How do we know we're not in a simulation right now?"

The simulation idea does more than illustrate the problem, however. It also *sharpens* the problem by turning Descartes's far-fetched scenarios involving evil demons into more realistic scenarios involving computers— scenarios we have to take seriously. In chapter 4, I make the case that the simulation idea undercuts many common responses to Descartes. In chapter 5, I use statistical reasoning about simulations to argue that we cannot know we're not in a simulation. All this makes Descartes's problem even harder.

Most importantly, reflection on virtual reality technology can help us *respond* to the problem of the external world. In chapters 6 through 9, I argue that if indeed we're in a simulation, tables and chairs are not illusions but perfectly real objects: they are digital objects that are made of bits. This leads us to what is sometimes called, in modern physics, the *it-from-bit hypothesis*: Physical objects are real and they are digital. Thinking about the simulation hypothesis and the it-from-bit hypothesis—two ideas inspired by modern computers—yields the beginnings of a response to Descartes's classic problem.

We can put Descartes's argument as follows: We don't know that we're not in a virtual world, and in a virtual world nothing is real, so we don't know that anything is real. This argument turns on the assumption that virtual worlds are not genuine realities. Once we make the case that virtual worlds are indeed genuine realities—and especially that objects in a virtual world are real—we can respond to Descartes's argument.

I shouldn't overstate the case. My analysis doesn't address everything Descartes says, and it doesn't prove that we know a great deal about the external world. Still, if the analysis works, it dissolves what is perhaps the Western tradition's prime reason for doubting that we can know anything about the external world. So it gives us at least a foothold in establishing that we have knowledge of the reality around us. …

Chapter 1
Is This the Real Life?

In the opening lines of the 1975 hit "Bohemian Rhapsody" by the British rock group Queen, lead singer Freddie Mercury sings in five-part harmony:

> *Is this the real life?*
> *Is this just fantasy?*

These questions have a history. Three of the great ancient traditions of philosophy—those of China, Greece, and India—all ask versions of Mercury's questions.

Their questions involve alternative versions of reality. Is this real life, or is it just a dream? Is this real life, or is it just an illusion? Is this real life, or is it just a shadow of reality?

Today we might ask: Is this real life, or is it virtual reality? We can think of dreams, illusions, and shadows as ancient counterparts of virtual worlds —minus the computer, which would not be invented for two millennia.

With or without the computer, these scenarios raise some of the deepest questions in philosophy. We can use them to introduce these questions and to guide our thinking about virtual worlds.

Zhuangzi's butterfly dream

The ancient Chinese philosopher Zhuangzi (also known as Zhuang Zhou or Chuang Tzu) lived around 300 BCE and was a central figure in the Daoist tradition. He recounts this famous parable: "Zhuangzi Dreams of Being a Butterfly."

> Once Zhuangzi dreamt he was a butterfly, a butterfly flitting and fluttering around, happy with himself and doing as he pleased. He didn't know he was Zhuangzi. Suddenly he woke up and there he was, solid and unmistakably Zhuangzi. But he didn't know if he was Zhuangzi who had dreamt he was a butterfly, or a butterfly dreaming he was Zhuangzi.

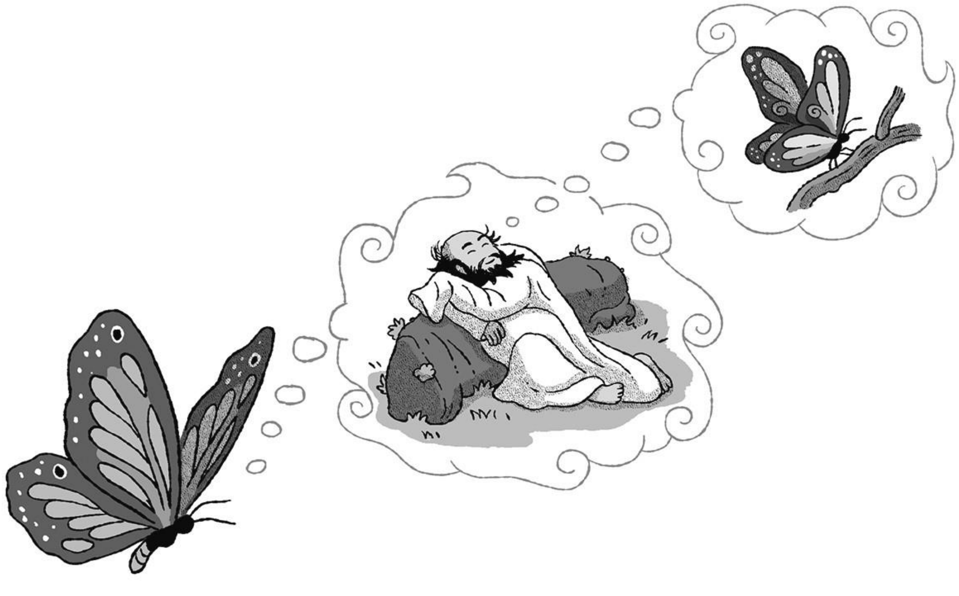

Figure 1 Zhuangzi's butterfly dream. Was he Zhuangzi who dreamt he was a butterfly, or a butterfly dreaming he was Zhuangzi?

Zhuangzi can't be sure that the life he's experiencing as Zhuangzi is real. Maybe the butterfly was real, and Zhuangzi is a dream.

A dream world is a sort of virtual world without a computer. So Zhuangzi's hypothesis that he is in a dream world right now is a computerfree version of the hypothesis that he's in a virtual world right now.

The plot of the Wachowski sisters' 1999 movie *The Matrix* provides a nice parallel. The main character, Neo, lives an ordinary life until he takes a red pill and wakes up in another world, where he's told that the world he knew was a simulation. If Neo had thought as deeply as Zhuangzi, he might have wondered, "Maybe my old life was the reality, and my new life is the simulation"—a perfectly reasonable thought. While his old world was a world of drudgery, his new world is a world of battles and adventure, where he's treated as a savior. Maybe the red pill knocked him out just long enough for him to be hooked up to this exciting simulation.

On one interpretation, Zhuangzi's butterfly dream raises a question about knowledge: How do any of us know we aren't dreaming right now? This is a cousin of the question raised in the introduction: How do any of us know we aren't in a virtual world right now? These questions lead to a more basic question: How do we know anything we experience is real?

Narada's transformation

Ancient Indian philosophers in the Hindu tradition were gripped by issues of illusion and reality. A central motif appears in the folk tale of the sage Narada's transformation. In one version of the story, Narada says to the god Vishnu, "I have conquered illusion." Vishnu promises to show Narada the true power of illusion (or *Maya*). Narada wakes up as a woman, Sushila, with no memory of what came before. Sushila marries a king, becomes pregnant, and eventually has eight sons and many grandsons. One day, an enemy attacks, and all her sons and grandsons are killed. As the queen grieves, Vishnu appears and says, "Why are you so sad? This is just an illusion." Narada finds himself back in his original body only a moment after the original conversation. He concludes that his whole life is an illusion, just like his life as Sushila.

Narada's life as Sushila is akin to life in a virtual world—a simulation with Vishnu acting as the simulator. As a simulator, Vishnu is in effect suggesting that Narada's ordinary world is a virtual world too.

Narada's transformation is echoed in an episode of the animated TV series *Rick and Morty,* which chronicles the interdimensional adventures of a powerful scientist, Rick, and his grandson Morty. Morty puts on a virtual reality helmet to play a video game titled *Roy: A Life Well Lived*. (It would be even better if Morty had played *Sue: A Life Well Lived*, but you can't have everything.) Morty lives Roy's entire fifty-five-year life: childhood, football star, carpet salesman, cancer patient, death. When he emerges from the game a moment later as Morty, his grandfather berates him for having made the wrong life decisions in the simulation. This is a recurring theme in the series. Its characters are in apparently normal situations that turn out to be simulations and are often led to ask whether their current reality might be a simulation, too.

Figure 2 Vishnu oversees Narada’s transformation into Sushila, in the style of *Rick and Morty*.

Narada’s transformation raises deep questions about reality. Is Narada’s life as Sushila real, or is it an illusion? Vishnu says it is an illusion, but this is far from obvious. We can raise an analogous question about virtual worlds, including the world of *Roy: A Life Well Lived*. Are these worlds real or illusory? An even more pressing question looms. Vishnu says that our ordinary lives are as illusory as Narada’s transformed life. Is our own world real or an illusion?

Plato’s cave

Around the same time as Zhuangzi, the ancient Greek philosopher Plato put forward his allegory of the cave. In his extended dialogue, the *Republic*, he tells the story of humans who are chained up in a cave, seeing only shadows cast on a wall by puppets that imitate things in the world of sunlight outside. The shadows are all the cave people know, so they take them to be reality. One day, one of them escapes and discovers the glories of the real world outside the cave. Eventually he reenters the cave and tells stories of that world, but no one believes him.

Plato’s prisoners watching shadows call to mind viewers in a movie theater. It’s as if the prisoners had never watched anything but movies—or, to update the technology, had watched only movies on a virtual reality headset. A 2016 mobile technology conference produced a famous photograph of Facebook chief executive Mark Zuckerberg walking down the aisle past the conference audience. The members of the audience are all wearing virtual reality headsets in the darkened hall, apparently unaware of Zuckerberg as he strides by. It's a contemporary illustration of Plato's cave.

Plato uses his allegory for many purposes. He's suggesting that our own imperfect reality is something like the cave. He's also using it to help us think about what sort of lives we want to live. In a key passage, Plato's spokesman, Socrates, raises the question of whether we should prefer life inside or outside the cave.

> SOCRATES: Do you think the one who had gotten out of the cave would still envy those within the cave and would want to compete with them who are esteemed and who have power? Or would not he much rather wish for the condition that Homer speaks of, namely "to live on the land [above ground] as the paid menial of another destitute peasant"? Wouldn't he prefer to put up with absolutely anything else rather than associate with those opinions that hold in the cave and be that kind of human being?
>
> GLAUCON: I think that he would prefer to endure everything rather than be that kind of human being.

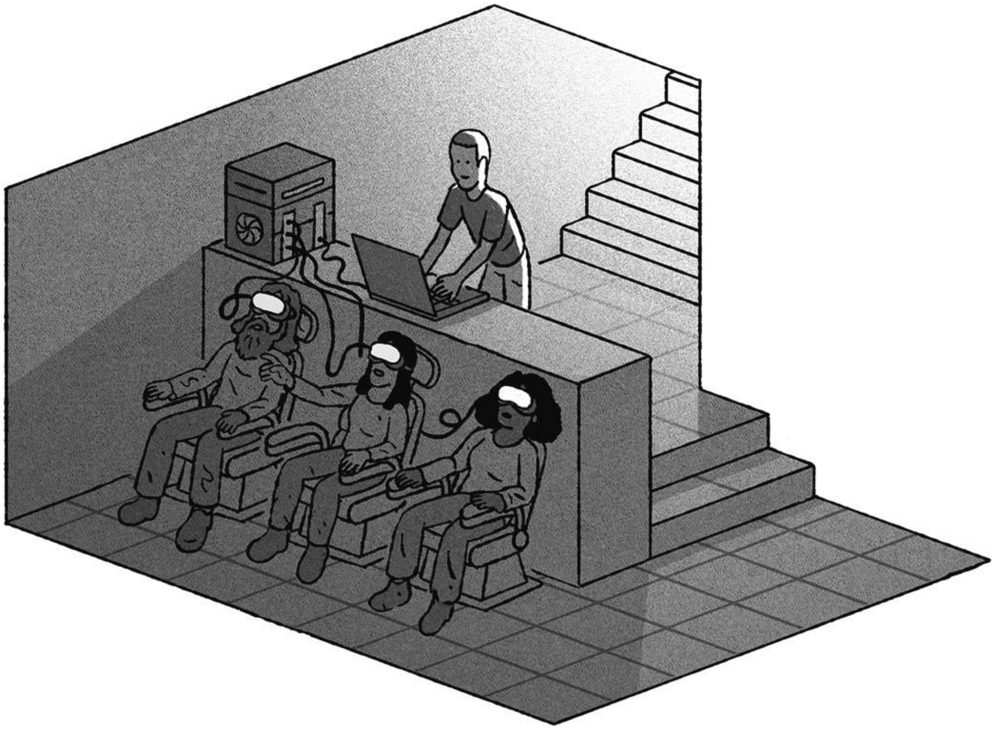

Figure 3 Plato's cave in the 21st century.

The allegory of the cave raises deep questions about value: that is, about good and bad, or at least about better and worse. Which is better, life inside the cave or life outside the cave? Plato's answer is clear: Life outside the cave, even life as a menial laborer, is vastly better than life inside it. We can ask the same question about virtual worlds. Which is better, life in a virtual world or life outside it? This leads us to a more fundamental question: What does it mean to live a good life?

Three questions

In one traditional picture, philosophy is the study of *knowledge* (How do we know about the world?), *reality* (What is the nature of the world?), and *value* (What is the difference between good and bad?).

Our three stories raise questions in each of these domains. Knowledge: *How can Zhuangzi know whether or not he's dreaming?* Reality: *Is Narada's transformation real or illusory?* Value: *Can one lead a good life in Plato's cave?*

When we transpose our three stories from their original realms of dreams, transformations, and shadows into the realm of virtuality, they raise three key questions about virtual worlds.

The first question, raised by Zhuangzi's butterfly dream, concerns knowledge. I'll call it the Knowledge Question. *Can we know whether or not we're in a virtual world?*

The second question, raised by Narada's transformation, concerns reality. I'll call it the Reality Question. *Are virtual worlds real or illusory?*

The third question, raised by Plato's cave, concerns value. I'll call it the Value Question. *Can you lead a good life in a virtual world?*

These three questions in turn lead us to three more general questions that are at the heart of philosophy: *Can we know anything about the world around us? Is our world real or illusory? What is it to lead a good life?*

Over the course of this book, these questions about knowledge, reality, and value will be at the heart of our exploration of virtual worlds and at the heart of our exploration of philosophy.

The Knowledge Question: Can we know whether or not we're in a virtual world?

In the 1990 movie *Total Recall* (remade with a few changes in 2012), the viewer is never quite sure which parts of the movie take place in a virtual world and which take place in the ordinary world. The main character, a construction worker named Douglas Quaid (played by Arnold Schwarzenegger) experiences many outlandish adventures on Earth and on Mars. At the movie's end, Quaid looks out over the surface of Mars and begins to wonder (and so do we) whether his adventures took place in the ordinary world or in virtual reality. The movie hints that Quaid may indeed be in a virtual world. Virtual reality technology that implants memories of adventures plays a key role in the plot. Since heroic adventures on Mars are presumably more likely to take place in virtual worlds than in ordinary life, Quaid, if he is reflective, will conclude that he's probably in virtual reality.

What about you? Can you know whether you're in a virtual or a nonvirtual world? Your life may not be as exciting as Quaid's. But the fact that you're reading a book about virtual worlds should give you pause. (The fact that I'm writing one should give me even more pause.) Why? I suspect that as simulation technology develops, simulators may be drawn to simulate people thinking about simulations, perhaps to see how close they come to realizing the truth about their lives. Even if we seem to be leading perfectly ordinary lives, is there any way we could know whether these lives are virtual?

To put my cards on the table: I don't know whether we're in a virtual world or not. I don't think you know, either. In fact, I don't think we can ever know whether or not we're in a virtual world. In principle, we could confirm that we *are* in a virtual world—for example, the simulators could choose to reveal themselves to us and show us how the simulation works. But if we're *not* in a virtual world, we'll never know that for sure.

… The basic reason is spelled out in chapter 2: We can never prove we're not in a computer simulation because any evidence of ordinary reality— whether the grandeur of nature, the antics of your cat, or the behavior of other people—could presumably be simulated.

… Going beyond this, we should take seriously the possibility that we *are* in a virtual world. The Swedish-born philosopher Nick Bostrom has argued on statistical grounds that under certain assumptions, there will be many more simulated people in the universe than nonsimulated people. If that's right, perhaps we should consider it likely that we're in a simulation. …

This verdict has major consequences for Descartes's problem: How do we know anything about the external world? If we don't know whether or not we're in a virtual world, and if nothing in a virtual world is real, then it looks like we cannot know if anything in the external world is real. And then it looks like we can't know anything at all about the external world. That's a shocking consequence. We can't know whether Paris is in France? I can't know that I was born in Australia? I can't know that there's a desk in front of me?

Central philosophical questions

To recap, our three main questions about virtual worlds are the following. The Reality Question*: Are virtual worlds real?* The Knowledge Question: *Can we know whether or not we're in a virtual world?* The Value Question: *Can you lead a good life in a virtual world?*

The Reality Question, the Knowledge Question, and the Value Question match up with three of the central divisions of philosophy.

(1) *Metaphysics*, the study of reality. Metaphysics asks questions like "What is the nature of reality?"
(2) *Epistemology*, the study of knowledge. Epistemology asks questions like "How can we know about the world?"
(3) *Value theory*, the study of values. Value theory asks questions like "What is the difference between good and bad?"

Or, to simplify: *What is this?* That's metaphysics. *How do you know?* That's epistemology. *Is it any good?* That's value theory.

When we ask the Reality Question, the Knowledge Question, and the Value Question, we're doing the metaphysics, epistemology, and value theory of virtual worlds. …

Answering philosophical questions

Philosophers are good at asking questions. We're less good at answering them. In 2020, my colleague David Bourget and I conducted a survey of around two thousand professional philosophers on one hundred central philosophical questions. To no one's surprise, we found large disagreement on the answers to almost all of them.

Every now and then a philosopher answers a question. Isaac Newton considered himself a philosopher. He worked on philosophical questions about space and time. He figured out how to answer some of them. As a result the new science of physics emerged. Something similar happened later with economics, sociology, psychology, modern logic, formal semantics, and more. All were founded or cofounded by philosophers who got clear enough on some central questions to help spin off a new discipline.

In effect, philosophy is an incubator for other disciplines. When philosophers figure out a method for rigorously addressing a philosophical question, we spin that method off and call it a new field. Because philosophy has been so successful at this over the centuries, what's now left in philosophy is a basket of hard questions that people are still figuring out. That's why philosophers disagree as much as they do.

Still, we can at least pose the questions and try our best to answer them. Occasionally a question is ready to be answered, and we'll get lucky. If we don't answer it, there's often value in the attempt. At the least, posing a question and exploring potential answers can lead us to understand the subject matter better. Others can build on that understanding, and eventually the question might be answered properly. …

Chapter 3
Do We Know Things?

The master argument for skepticism

Philosophers love arguments. This is not to say that they love disputes with each other, although many enjoy that, too. In philosophy, an argument is a chain of reasoning that supports a conclusion. I can argue that God exists by laying out some reasons for thinking that God exists and showing how they support my conclusion.

Sometimes arguments are informal. Maybe I try to convince you that we should go to a movie by giving some reasons: We both have spare time, it's a great movie, and it's only playing tonight. I can do the same in philosophy. I can try to convince you that you can't be certain of the world around you by giving some reasons: You've had sensory illusions before, so how do you know you're not having one now? If I do a good job, maybe it will convince you of the conclusion, or at least prompt you to take it seriously.

Sometimes arguments are formal. That may sound intimidating, but formal arguments are often simple. You lay out a number of claims that are *premises*, and then you lay out a *conclusion* that follows from them. Usually the idea is that the premises are plausible enough that people will have some inclination to accept them, and the conclusion drawn from these premises is bold enough to be interesting.

Here's a formal argument for skepticism about the external world.

1. You can't know you're not in a simulation.
2. If you can't know you're not in a simulation, you can't know anything about the external world.

---

3. So: You can't know anything about the external world.

Here, the first two claims are the premises, and the third claim is the conclusion. The conclusion follows logically from the premises: If the premises are true, the conclusion has to be true. When the conclusion follows from the premises, philosophers say the argument is *valid*. When in addition the premises are true, the argument is *sound*. Just because an argument is valid, this doesn't mean that the conclusion is true. After all, one or both of the premises could be false. But when an argument is sound, the conclusion has to be true. In the argument above, *if* you accept the two premises, you pretty much have to accept the conclusion.

Bertrand Russell once said, "The point of philosophy is to start with something so simple as not to seem worth stating, and to end with something so paradoxical that no one will believe it." The argument above at least has the potential to meet Russell's ideal. Both premises seem plausible, at least on a moment's reflection, and the conclusion seems surprising. That’s one of the things that makes this argument so interesting.

In fact, this argument is so interesting that it, or something like it, is often regarded as the *master argument* for skepticism in recent philosophy. The details can change a bit. For example, we could replace simulations with evil demons or brains in vats, but the basic idea is intact.

Why believe the first premise? … In a good-enough simulation, the world would look and feel to you exactly as today’s world looks and feels to you now. And if a simulation would look and feel the same as reality, it’s hard to see how we could know we’re in a simulation rather than reality.

Why believe the second premise? Pick anything you think you knew about the external world. You thought you knew that Paris is in France, or that there’s a spoon in front of you. But if you’re in a simulation, then your beliefs about Paris and the spoon come from the simulation, not from reality. Paris and the spoon are simulated. The world outside the simulation may be entirely different. There may well be no Paris and no spoon in reality outside the simulation. So to know that Paris is in France or that there’s truly a spoon in front of you, you have to rule out the possibility that you’re in a simulation.

The reasoning here is a bit like this: If your phone is a knockoff, you don’t really have an iPhone. So if you can’t know that your phone isn’t a knockoff, you can’t know that you have an iPhone. In this case, we start from the plausible claim: If you’re in a simulation, there’s no spoon in front of you. By the same sort of reasoning as in the iPhone case, we get to: If you can’t know you’re not in a simulation, you can’t know there’s a spoon in front of you. The same applies to everything in the external world.

Our Reality Question about virtual reality was: *Is virtual reality real or an illusion?* If you answer by saying *Virtual reality is an illusion*, you’ll probably accept the second premise. Here’s why. Given this answer, you’ll also accept *Simulations are illusions*, since simulations are a kind of virtual reality in the broad sense. In fact, you’ll probably accept *If you’re in a simulation, everything you experience in the external world is illusory*. So if you can’t rule out the simulation hypothesis, you can’t rule out that everything in the external world is illusory. It seems to follow that you can’t know anything about the external world at all.

The conclusion is startling. If you’re like most people, you thought you knew a lot of things. You thought you knew that Paris is in France, and you thought you knew what’s physically in front of you. But it turns out you don’t! The argument applies to more than just objects or cities. It applies to memories of your childhood. If you’re in a simulation, so the argument goes, your memories of going to school aren’t real, so you don’t really know that you went to school. The same goes for pretty much everything you thought you knew about the external world and your life in it.

Strictly speaking, the argument doesn't stop you from knowing a *few* things about the external world. Some things are true as a matter of logic or mathematics. You can know that all dogs are dogs, for example. You can know that if there is one table here and a different table there, there are two tables. But these are all trivialities. To be strictly correct, we could adjust the conclusion to "We can't know anything substantial about the external world."

If we accept the premises, the argument leads us to global skepticism about the external world—that is, the view that we don't know anything substantial about the external world. Maybe we can still know that two plus two is four, but that's not a huge consolation.

What can we do to avoid the shocking conclusion?

I think, therefore I am

Descartes himself didn't want to be a skeptic. In fact, he wanted to establish a foundation for all knowledge. So after casting all our knowledge into doubt with his skeptical arguments, he tried to build it back up, piece by piece.

Descartes needed to start with a piece of knowledge he couldn't doubt. He needed to uncover something about reality that would be true even if he was having sensory illusions, even if he was dreaming, even if he was being fooled by an evil demon. He found a candidate: his own existence.

Descartes's famous argument for his own existence, presented most explicitly in his 1637 *Discourse on Method*, went like this: *Cogito, ergo sum*. I think, therefore I am.

Philosophers have interpreted Descartes's celebrated slogan in many different ways. But at least on the surface, it looks like an argument. The premise of the argument (to unpack it a little) is *I am thinking*. The conclusion is *I exist*. As with most arguments, the real work is done by the premise. Once you grant that, the conclusion *I exist* seems to follow as a matter of logic.

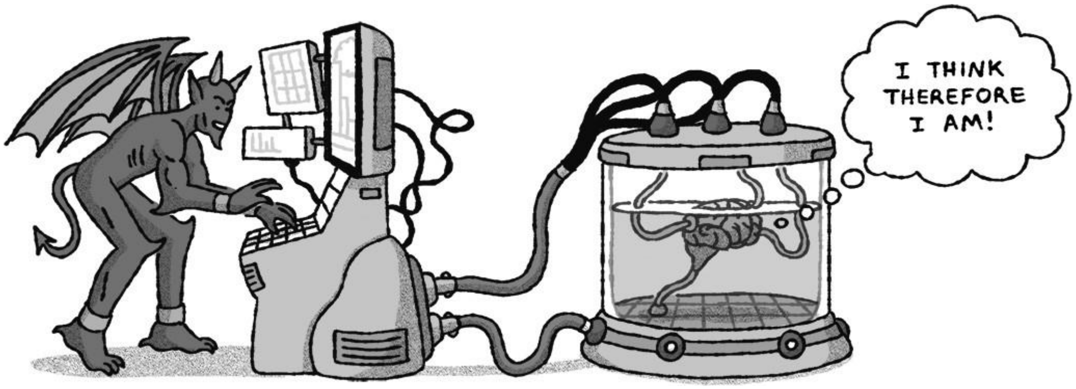

Figure 9 Even if you're a brain in a vat, receiving sensations from an evil demon, you can still reason, "I think, therefore I am."

How does Descartes know he's thinking? For a start, this knowledge does not seem to be undercut by the skeptical arguments. Even if you're in the grip of a sensory illusion, you're still thinking. Even if you're dreaming, you're still thinking. Even if you're being fooled by an evil demon, you're still thinking. Even if you're a brain in a vat, you're still thinking. Even if you're in a simulation, you're still thinking.

More deeply, Descartes reasoned that he could not doubt that he's thinking. Even if he doubted that he was thinking, his doubt was itself a sort of thinking. To doubt that one is thinking is internally inconsistent: The doubting itself shows that the doubt is wrong.

Once Descartes knew he was thinking, it was a small step to knowing his own existence. Where there is thinking, there must be a thinker. So Descartes concludes: *Sum*! I exist!

… [A] lot of people accept Descartes's *Cogito, ergo sum*. It's hard to doubt that I'm thinking. The evil-demon scenario doesn't really call my own mind into doubt, and it's not easy to generate scenarios that do. As a result, even some skeptical philosophers are prepared to say that we do know that we think, and that therefore we do know that we exist.

Speaking for myself, I don't think there's anything special about thinking per se. Descartes could have said, "I feel, therefore I am," or "I see, therefore I am," or "I worry, therefore I am." All of these are claims about his mind that he can be certain of and that aren't threatened by the evil demon. At least, he can be certain about these claims if they're understood as states of consciousness, or subjective experience. If we understand "see" as referring simply to the subjective experience of seeing, then Descartes can be certain he is seeing. …

If we grant *Cogito, ergo sum*, that gives Descartes a foundation. The hard part is what comes next. How do we get from knowledge of ourselves and our own minds to knowledge of the external world? …