



# Contents

---

*Introduction* Adventures in technophilosophy **xi**

## **Part 1 VIRTUAL WORLDS**

*Chapter 1* Is this the real life? **3**

*Chapter 2* What is the simulation hypothesis? **20**

## **Part 2 KNOWLEDGE**

*Chapter 3* Do we know things? **43**

*Chapter 4* Can we prove there is an external world? **61**

*Chapter 5* Is it likely that we're in a simulation? **81**

## **Part 3 REALITY**

*Chapter 6* What is reality? **105**

*Chapter 7* Is God a hacker in the next universe up? **124**

*Chapter 8* Is the universe made of information? **145**

*Chapter 9* Did simulation create its from bits? **167**

## **Part 4 REAL VIRTUAL REALITY**

*Chapter 10* Do virtual reality headsets create reality? **185**

*Chapter 11* Are virtual reality devices illusion machines? **203**

*Chapter 12* Does augmented reality lead to alternative facts? **225**

*Chapter 13* Can we avoid being deceived by deepfakes? **238**

## ***Part 5* MIND**

*Chapter 14* How do mind and body interact in a virtual world? **255**

*Chapter 15* Can there be consciousness in a digital world? **274**

*Chapter 16* Does augmented reality extend the mind? **294**

## ***Part 6* VALUE**

*Chapter 17* Can you lead a good life in a virtual world? **311**

*Chapter 18* Do simulated lives matter? **331**

*Chapter 19* How should we build a virtual society? **350**

## ***Part 7* FOUNDATIONS**

*Chapter 20* What do our words mean in virtual worlds? **367**

*Chapter 21* Do dust clouds run computer programs? **385**

*Chapter 22* Is reality a mathematical structure? **399**

*Chapter 23* Have we fallen from the Garden of Eden? **423**

*Chapter 24* Are we Boltzmann brains in a dream world? **440**

*Acknowledgments* **463**

*Glossary* **467**

*Notes* **471**

*Index* **505**

## Introduction

# Adventures in technophilosophy

**W**HEN 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.



The central thesis of this book is: *Virtual reality is genuine reality*. Or at least, *virtual realities are genuine realities*. Virtual worlds need not be second-class realities. They can be first-class realities.

We can break down this thesis into three parts:

- Virtual worlds are not illusions or fictions, or at least they need not be. What happens in VR really happens. The objects we interact with in VR are real.
- Life in virtual worlds can be as good, in principle, as life outside virtual worlds. You can lead a fully meaningful life in a virtual world.
- The world we're living in could be a virtual world. I'm not saying it is. But it's a possibility we can't rule out.

The thesis—especially the first two parts—has practical consequences for the role of VR technology in our lives. In principle, VR can be much more than escapism. It can be a full-blooded environment for living a genuine life.

I'm not saying that virtual worlds will be some sort of utopia. Like the internet, VR technology will almost certainly lead to awful things as well as wonderful things. It's certain to be abused. Physical reality is abused, too. Like physical reality, virtual reality has room for the full range of the human condition—the good, the bad, and the ugly.

I'll focus more on VR in principle than VR in practice. In practice, the road to full-scale virtual reality is sure to be bumpy. It won't surprise me if widespread adoption of VR is limited for a decade or two, while the technology matures. No doubt it will move in all sorts of directions I haven't anticipated. But once a mature VR technology is developed, it should be able to support lives that are on a par with or even surpass life in physical reality.



The title of this book captures my main claims. You can understand it in a number of ways. Each virtual world is a new reality: Reality+. Augmented reality involves additions to reality: Reality+. Some virtual worlds are as good as or better than ordinary reality: Reality+. If we're in a simulation, there is more to reality than we thought: Reality+. There will be a smorgasbord of multiple realities: Reality+.

I know that what I'm saying is counterintuitive to many people. Perhaps you think that VR is Reality-, or Reality Minus. Virtual worlds are fake realities, not genuine realities. No virtual world is as good as ordinary reality. Over the course of this book, I'll try to convince you that Reality+ is closer to the truth.



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.

We’ll also use technology to illuminate traditional questions about the mind: How do mind and body interact? (See chapter 14.) What is consciousness? (See chapter 15.) Does the mind extend beyond the body? (See chapter 16.) In each case, thinking about a technology—VR, artificial intelligence (AI), and augmented reality (AR), respectively—can illuminate those questions. And conversely, thinking about the questions can illuminate these technologies.

It’s worth saying that my views about consciousness and the mind are not the main focus of this book. I’ve explored those issues in other work, and this book is independent of them to a large degree. I hope that even people who disagree with me about consciousness may find my picture of reality appealing. That said, there are many connections between the two domains. You can think of chapters 15 and 16, in particular, as adding a fourth plank to the thesis that virtual reality is genuine reality: namely, *virtual and augmented minds are genuine minds*.

Technology can also illuminate traditional questions about value and ethics. Value is the domain of good and bad, better and worse. Ethics is the domain of right and wrong. What makes for a good life? (See chapter 17.) What is the difference between right and wrong? (See chapter 18.) How should society be organized? (See chapter 19.) I’m by no means an expert on these issues, but technology provides at least an interesting angle on them.

Other time-hallowed philosophical questions will come up along the way. Is there a God? (See chapter 7.) What is the universe made of? (See chapter 8.) How does language describe reality? (See chapter

20.) What does science tell us about reality? (See chapters 22 and 23.) It turns out that to make our case that virtual reality is genuine reality, we have to think hard about those old questions. As always, the illumination flows both ways; thinking about technology throws light on the old questions in turn.



I also want to use philosophy to address new questions about technology, especially the technology of virtual worlds. These include questions about everything from video games through augmented reality glasses and virtual reality headsets to simulations of entire universes.

I've already outlined my central thesis that virtual reality is genuine reality. Where VR is concerned, I'll ask questions like: Is virtual reality an illusion? (See chapters 6, 10, and 11.) What are virtual objects? (See chapter 10.) Does augmented reality genuinely augment reality? (See chapter 12.) Can you live a good life in VR? (See chapter 17.) How should you behave in a virtual world? (See chapter 19.)

I'll also discuss other technologies: artificial intelligence, smartphones, the internet, deepfakes, and computers in general. How can we know we're not being deceived by deepfakes? (See chapter 13.) Can AI systems be conscious? (See chapter 15.) Do smartphones extend our minds, and is the internet making us smart or stupid? (See chapter 16.) And what is a computer, anyway? (See chapter 21.)

These questions are all philosophical questions. Many of them are also intensely practical questions. We need to make decisions right now about how we use video games, smartphones, and the internet. An increasing number of such practical questions will confront us in decades to come. As we spend more and more time in virtual worlds, we'll have to grapple with the issue of whether life there is fully meaningful. Eventually, we may have to decide whether or not to upload ourselves to the cloud entirely. Thinking philosophically can help us get clear on these decisions about how to live our lives.



By the end of this book, you'll have been introduced to many of the central questions in philosophy. We'll encounter both historical greats from centuries and millennia past and contemporary figures and arguments from recent decades. We'll cover many of the central topics in philosophy: knowledge, reality, mind, language, value, ethics, science, religion, and more. I'll introduce some of the powerful tools that philosophers have developed over the centuries for thinking about these issues. This is only one perspective, and a great deal of important philosophy is left out. But by the end, you'll have a sense of some of the historical and contemporary landscape of philosophy.

To help readers think through these ideas, I've made connections to science fiction and other corners of popular culture whenever I can. Many authors of science fiction have delved into these issues just as deeply as philosophers have. I've often had new philosophical ideas by thinking about science fiction. Sometimes I think science fiction gets these issues right, and sometimes it gets them wrong. Either way, science-fiction scenarios can prompt a lot of fruitful philosophical analysis.

The best way I know to introduce philosophy is to *do* philosophy. So while I'll start many chapters by posing a philosophical question connected to virtual worlds and introducing some philosophical background, I'll usually get down quickly to thinking hard about the issues. I'll analyze the issues both inside and outside virtual worlds, with an eye on building my argument for the Reality+ point of view.

As a result, this book is as full of my own philosophical theses and arguments as anything I've ever written. While some chapters of the book go over ground I've discussed in academic articles, well over half of it is entirely new. So even if you're an old hand at philosophy, I hope that you'll find rewards here. In an online supplement ([consc.net/reality](http://consc.net/reality)), I've included extensive notes and appendices pursuing the issues in more depth, often including connections to the academic literature.



The book has seven parts. Part 1 (chapters 1 and 2) introduces the central problems of the book and the simulation hypothesis that plays a

central role. Part 2 (chapters 3–5) focuses on questions about knowledge, and especially Descartes’s arguments for skepticism about the external world. Part 3 (chapters 6–9) focuses on questions about reality, and makes an initial case for my thesis that virtual reality is genuine reality.

The next three parts of the book develop many different aspects of the thesis. Part 4 (chapters 10–13) brings things down to earth to focus on questions about real virtual reality technology: virtual reality headsets, augmented reality glasses, and deepfakes. Part 5 (chapters 14–16) focuses on questions about the mind. Part 6 (chapters 17–19) focuses on questions about value and ethics. Finally, part 7 (chapters 20–24) focuses on foundational issues about language, computers, and science that are required to fully develop the Reality+ vision. The last chapter pulls the pieces together to see where things stand with Descartes’s problem of the external world.

Different readers may want to read the book in different ways. Everyone should read chapter 1, but after that you can strike out in many different directions. In the endnotes, I give some possible paths, depending on your interests. Many chapters stand relatively independently. Chapters 2, 3, 6, and 10 may be especially helpful in providing background for the chapters that follow, but they aren’t absolutely essential.

Most of the chapters are frontloaded with introductory material toward the start. The discussion sometimes gets denser toward the end of each chapter, and toward the end of the book. If you’re after a shorter book and a lighter reading experience, you might try reading the first two or three sections of every chapter, and then skipping to the next chapter whenever you like.



We live in an age in which truth and reality have been under attack. We’re sometimes said to be in an era of post-truth politics in which truth is irrelevant. It’s common to hear that there’s no absolute truth and no objective reality. Some people think that reality is all in the mind, so that what’s real is entirely up to us. The multiple realities of

this book may initially suggest a view like that on which truth and reality are cheap. This is not my view.

Here's my view of these things. Our minds are part of reality, but there's a great deal of reality outside our minds. Reality contains our world and it may contain many others. We can build new worlds and new parts of reality. We know a little about reality, and we can try to know more. There may be parts of it that we can never know.

Most importantly: Reality exists, independently of us. The truth matters. There are truths about reality, and we can try to find them. Even in an age of multiple realities, I still believe in objective reality.

Part 1

---

# **VIRTUAL WORLDS**



## 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.

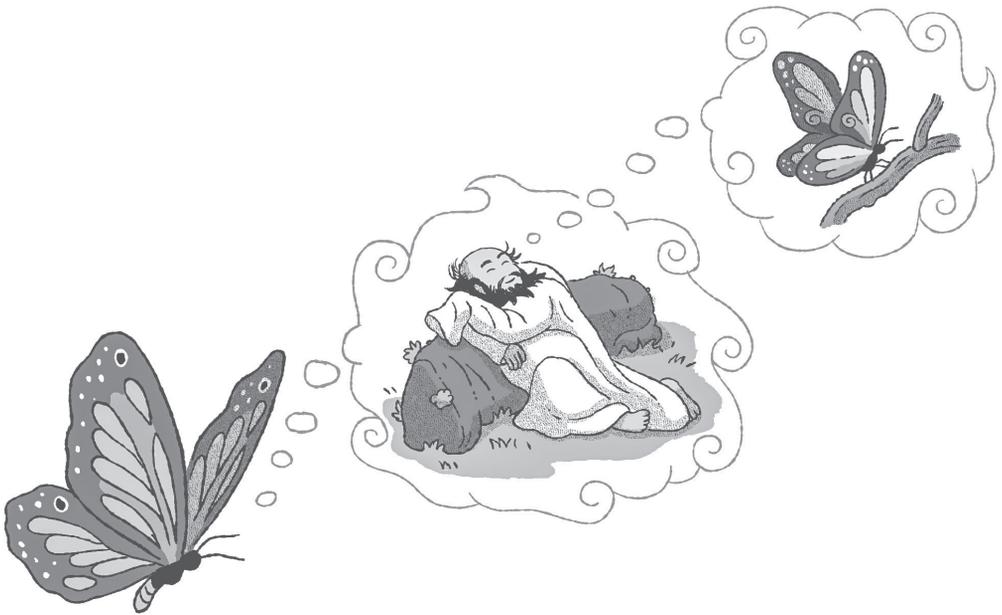
With or without the computer, these scenarios raise some of the deepest questions in philosophy. We can use them to introduce these questions and to guide our thinking about virtual worlds.

## 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.

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.



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

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 computer-free 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 conversa-



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

tion. 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.

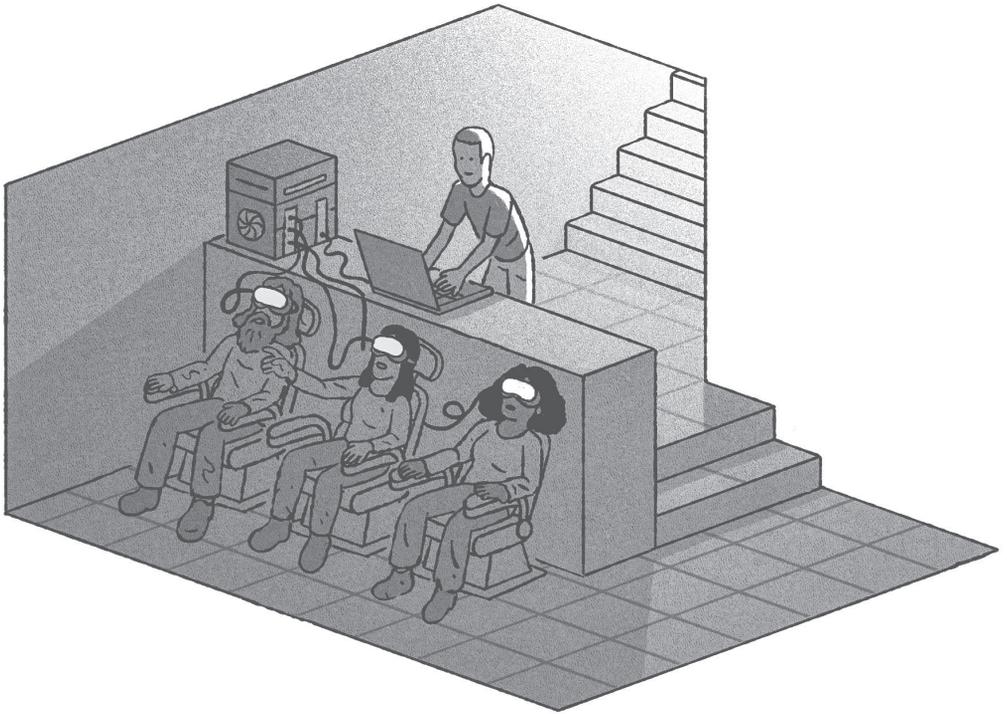
Narada's transformation raises deep questions about reality. Is Narada's life as Sushila real, or is it an illusion? Vishnu says it is an illusion, but this is far from obvious. We can raise an analogous question about virtual worlds, including the world of *Roy: A Life Well Lived*. Are these worlds 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,



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

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.

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, real-

ity, 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 non-virtual 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.

I'll discuss the reasons for this uncertainty over the next few chapters. 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.

Over the centuries, many philosophers have offered strategies that could be used to show that we're not in a virtual world. I'll discuss these strategies in chapter 4 and argue that they don't work. 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. I'll argue in chapter 5 for a somewhat weaker conclusion: All these considerations mean that we can't know we're *not* 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?

Many philosophers try to avoid this shocking consequence by arguing for a positive answer to the Knowledge Question: we *can* know that we're not in a simulation. If we can know that, then we can know something about the external world after all. If I'm right, though, we can't fall back on this comforting position. We can't know that we're not in a simulation. That makes the problem of knowledge of the external world that much harder.

## The Reality Question: Are virtual worlds real or illusory?

Whenever virtual reality is discussed, one hears the same refrain. *Simulations are illusions. Virtual worlds aren't real. Virtual objects don't really exist. Virtual reality isn't genuine reality.*

You can find this idea in *The Matrix*. In a waiting room inside the simulation, Neo sees a child apparently bending a spoon with the power of his mind. They engage in conversation:

CHILD: Do not try and bend the spoon. That's impossible.

Instead . . . only try to realize the truth.

NEO: What truth?

CHILD: There is no spoon.

This is presented as a deep truth. *There is no spoon.* The spoon inside the Matrix is not real but a mere illusion. The implication is that everything one experiences in the Matrix is an illusion.

In a commentary on *The Matrix*, the American philosopher Cornel West, who himself played Councillor West of Zion in *The Matrix Reloaded* and *The Matrix Revolutions*, takes this line of thinking a step further. Speaking of awakening from the Matrix, he says “What you think you're awakening to may in fact be another species of illusion. It's illusions all the way down.” Here there is an echo of Vishnu: Simulations are illusions, and ordinary reality may be an illusion, too.

The same line of thinking recurs in the TV series *Atlanta*. Three characters are sitting around a pool late at night discussing the simulation hypothesis. Nadine becomes convinced: “We're all nothing. It's a simulation, Van. We're all fake.” She takes for granted that if we're living in a simulation, we're not real.

I think these claims are wrong. Here's what I think: *Simulations are not illusions. Virtual worlds are real. Virtual objects really exist.* In my view, the Matrix child should have said, “Try to realize the truth. There is a spoon—a digital spoon.” Neo's world is perfectly real. So is Nadine's world, even if she is in a simulation.

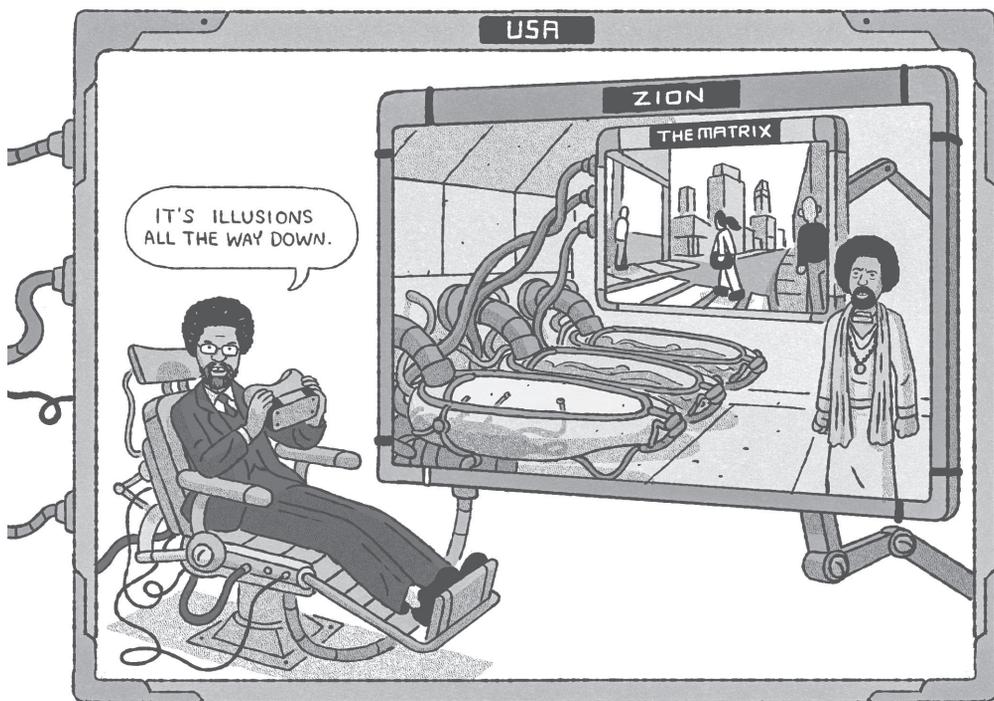


Figure 4 Cornel West, awakening from life as Councillor West of Zion, on illusion and reality.

The same goes for our world. Even if we're in a simulation, our world is real. There are still tables and chairs and people here. There are cities, there are mountains, there are oceans. Of course there may be many illusions in our world. We can be deceived by our senses and by other people. But the ordinary objects around us are real.

What do I mean by “real”? That’s complicated—the word “real” doesn’t have a single, fixed meaning. In chapter 6, I’ll discuss five different criteria for being “real.” I’ll argue that even if we’re in a simulation, the things we perceive meet all these criteria for reality.

What about ordinary virtual reality, experienced through a headset? This can sometimes involve illusion. If you don’t know you’re in VR and you take the virtual objects to be normal physical objects, you’d be wrong. But I’ll argue in chapter 11 that for experienced users of VR,

who know they're using VR, there need be no illusion. They're experiencing real virtual objects in virtual reality.

Virtual realities are different from nonvirtual realities. Virtual furniture isn't the same as nonvirtual furniture. Virtual entities are made one way, and nonvirtual entities are made another. Virtual entities are *digital* entities, made of computational and informational processes. More succinctly, they're made of bits. They're perfectly real objects that are grounded in a pattern of bits in a computer. When you interact with a virtual sofa, you're interacting with a pattern of bits. The pattern of bits is entirely real, and so is the virtual sofa.

"Virtual reality" is sometimes taken to mean "fake reality." If I'm right, that's the wrong way to define it. Instead it means something closer to "digital reality." A virtual chair or table is made of digital processes, just as a physical chair or table is made of atoms and quarks and ultimately of quantum processes. The virtual object is different from the nonvirtual one, but both are equally real.

If I'm right, then Narada's life as a woman is not entirely an illusion. Nor is Morty's life as a football star and carpet salesman. The long lives that they experience really happen. Narada really lives a life as Sushila. Morty really lives a life as Roy, albeit in a virtual world.

This view has major consequences for the problem of the external world. If I'm right, then even if I don't know whether or not we're in a simulation, it won't follow that I don't know whether or not the objects around us are real. If we're in a simulation, tables are real (they're patterns of bits), and if we're not in a simulation, tables are real (they're something else). So either way, tables are real. This offers a new approach to the problem of the external world, one that I will spell out over the course of this book.

## **The Value Question: Can you live a good life in a virtual world?**

In James Gunn's 1954 science-fiction story "The Unhappy Man," a company known as Hedonics, Inc., uses the new "science of happiness" to

improve people's lives. People sign a contract to move their life into "sensies," a sort of virtual world where everything is perfect:

We take care of everything; we arrange your life so you never have to worry again. In this age of anxiety, you never have to be anxious. In this age of fear, you never need be afraid. You will always be fed, clothed, housed, and happy. You will love and be loved. Life, for you, will be an unmixed joy.

Gunn's protagonist rejects the offer to hand over his life to Hedonics, Inc.

In his 1974 book *Anarchy, State, and Utopia*, the American philosopher Robert Nozick offers the reader a similar choice:

Suppose there was an experience machine that would give you any experience you desired. Super-duper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life experiences?

Gunn's sensies and Nozick's experience machine are virtual reality devices of a kind. They are asking, "Given the choice, would you spend your life in this kind of engineered reality?"

Like Gunn's protagonist, Nozick says no, and he expects his readers to do the same. His view seems to be that the experience machine is a second-class reality. Inside the machine, one does not actually do the things one seems to be doing. One is not a genuine autonomous person. For Nozick, life in the experience machine does not have much meaning or value.

Many people would agree with Nozick. In a 2020 survey of professional philosophers, 13 percent of respondents said they would enter the experience machine, and 77 percent said they would not. In broader surveys, most people decline the opportunity, too—although as virtual

worlds have become more and more a part of our lives, the number who say they would plug in is increasing.

We can ask the same question of VR more generally. Given the chance to spend your life in VR, would you do it? Could this ever be a reasonable choice? Or we can ask the Value Question directly: Can you lead a valuable and meaningful life in VR?

Ordinary VR differs in some ways from Nozick's experience machine. You know when you're in VR, and many people can enter the same VR environment at once. In addition, ordinary VR is not entirely preprogrammed. In interactive virtual worlds, you make real choices rather than simply living out a script.

Still, in a 2000 article in *Forbes* magazine, Nozick extends his negative assessment of the experience machine to ordinary VR. He says: "even if everybody were plugged into the same virtual reality, that wouldn't be enough to make its contents truly real." He also says of VR: "The pleasures of this may be so great that many people will choose to spend most of their days and nights that way. Meanwhile, the rest of us are likely to find that choice deeply disturbing."

Where VR is concerned, I'll argue (in chapter 17) that Nozick's answer is the wrong answer. In full-scale VR, users will build their own lives as they choose, genuinely interacting with others around them and leading a meaningful and valuable life. Virtual reality need not be a second-class reality.

Even existing virtual worlds—such as *Second Life*, which has been perhaps the leading virtual world for building a day-to-day life since it was founded in 2003—can be highly valuable. Many people have meaningful relationships and activities in today's virtual worlds, although much that matters is missing: proper bodies, touch, eating and drinking, birth and death, and more. But many of these limitations will be overcome by the fully immersive VR of the future. In principle, life in VR can be as good or as bad as life in a corresponding nonvirtual reality.

Many of us already spend a great deal of time in virtual worlds. In the future, we may well face the option of spending more time there, or even of spending most of our lives there. If I'm right, this will be a reasonable choice.

Many would see this as a dystopia. I do not. Certainly virtual worlds can be dystopian, just as the physical world can be, but they won't be dystopian merely because they're virtual. As with most technologies, whether VR is good or bad depends entirely on how it's used.

## Central philosophical questions

To recap, our three main questions about virtual worlds are the following. The Reality Question: *Are virtual worlds real?* (My answer: yes.) The Knowledge Question: *Can we know whether or not we're in a virtual world?* (My answer: no.) The Value Question: *Can you lead a good life in a virtual world?* (My answer: yes.)

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.

Other philosophical questions we'll ask about virtual worlds include:

The Mind Question: *What is the place of minds in virtual worlds?*

The God Question: *If we're in a simulation, is there a god?*

The Ethics Question: *How should we act in a virtual world?*

The Politics Question: *How should we build a virtual society?*

The Science Question: *Is the simulation hypothesis a scientific hypothesis?*

The Language Question: *What is the meaning of language in a virtual world?*

Like our three main questions, these six further questions each correspond to an area of philosophy: the philosophy of mind, the philosophy of religion, ethics, political philosophy, the philosophy of science, and the philosophy of language.

The traditional questions in each of these areas are more general: What is the place of minds in reality? Is there a God? How should we treat other people? How should society be organized? What does science tell us about reality? What is the meaning of language?

In addressing the questions about virtual worlds, I'll do my best to connect them to these bigger questions, too. That way, our answers will not just help us come to grips with the role of virtual worlds in our lives. They'll also help us to get clearer on reality itself.

## 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 phi-

losophers 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.

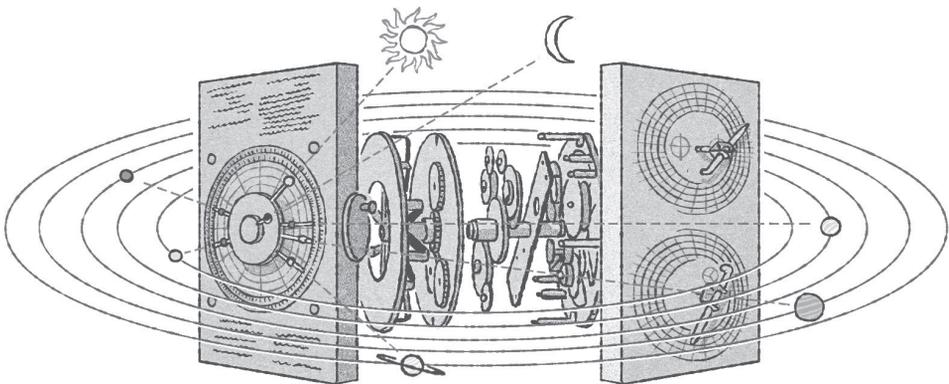
Still, we can at least pose the questions and try our best to answer them. Occasionally a question is ready to be answered, and we'll get lucky. If we don't answer it, there's often value in the attempt. At the least, posing a question and exploring potential answers can lead us to understand the subject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

In this book, I'll try to answer some of the questions I've posed. I can't expect you to agree with all of my answers. Still, I hope you might find understanding in the attempt. With luck, there will be something here that someone can build on. One way or another, we can hope that some of these questions about virtual worlds will eventually migrate from philosophy to a new discipline of their own.

## Chapter 2

# What is the simulation hypothesis?

**T**HE ANTIKYTHERA MECHANISM WAS FOUND IN A SHIPWRECK off the coast of the Greek island of Antikythera in 1901. It dates from two thousand years earlier. The mechanism is a bronze device that was originally mounted in a wooden box about 13 inches across. Superficially, it resembles a clock, with a complex system of 30 or more gears that once drove pointers and dials on the front and the back. Through painstaking analysis over the last century, researchers have discovered that the pointers simulate the day-by-day positions of the Sun and Moon in the zodiac according to the theories of the astronomer Hipparchus of Rhodes. Recently, mathematical analysis of surviving text and gear fragments has provided strong evidence that the system simulated the five known planets as well. It appears that the



*Figure 5* A reconstruction of the Antikythera mechanism, which simulated the position of the Sun and the Moon and probably the five known planets.

Antikythera mechanism is an attempt to simulate the solar system. It is the first known cosmic simulation.

The Antikythera mechanism is a *mechanical simulation*. In a mechanical simulation, the positions of components reflect the positions of the entities they're simulating. In the Antikythera, the motion of the gears is intended to reflect the motion of the Sun and Moon against the stars. One could use it to predict a solar eclipse years in the future.

Mechanical simulations are still used from time to time. One prominent example is a mechanical simulation of the San Francisco Bay and its environs, erected in a giant warehouse taking up more than an acre just outside San Francisco. It's a scale model, with enormous amounts of water moved by hydraulic mechanisms to simulate tides, currents, and other forces. It was built to test whether a plan for building dams on the bay would work. The mechanical simulation showed that it wouldn't, and the dams were never built.

Mechanical simulations of highly complex systems are difficult to build, and the art and science of simulation didn't flourish until the start of the computer age in the mid-20th century. In the celebrated code-breaking unit in Bletchley Park (depicted in the film *The Imitation Game*), the British mathematician Alan Turing and other researchers built some of the first computers in order to simulate and analyze German code systems. After the war, the mathematical physicists Stanislaw Ulam and John von Neumann used the ENIAC computer to simulate the behavior of neutrons in a nuclear explosion.

These models were among the first computer simulations. Whereas a mechanical simulation is driven by physical mechanisms, a computer simulation is driven by algorithms. Instead of using pointers and gears to reflect the positions of the planets, a modern computer simulation uses patterns of bits. An algorithmic simulation of the observed laws of planetary motion makes sure that the bits evolve in a way that reflects the positions of the planets. Using this method, we now have accurate simulations of the solar system allowing us to predict the position of Mars with uncanny precision.

Computer simulations are ubiquitous in science and engineering. In

physics and chemistry, we have simulations of atoms and molecules. In biology, we have simulations of cells and organisms. In neuroscience, we have simulations of neural networks. In engineering, we have simulations of cars, planes, bridges, and buildings. In planetary science, we have simulations of Earth's climate over many decades. In cosmology, we have simulations of the known universe as a whole.

In the social sphere, there are many computer simulations of human behavior. As early as 1955, Daniel Gerlough completed a PhD thesis on computer simulation of freeway traffic. In 1959, the Simulmatics Corporation was founded to simulate and predict how a political campaign's messaging would affect various groups of voters. It was said that this effort had a significant effect on the 1960 US presidential election. The claim may have been overblown, but since then, social and political simulations have become mainstream. Advertising companies, political consultants, social media companies, and social scientists build models and run simulations of human populations as a matter of course.

Simulation technology is improving fast, but it's far from perfect. A simulation usually concentrates on a certain level. A population-level simulation approximates human behavior with simple psychological models, but it doesn't usually try to simulate the neural networks that underlie the psychology. A hot topic in the science of simulation involves multiscale simulations, which are increasingly able to simulate systems at a number of levels simultaneously, but there are limits. There are no useful simulations of human behavior that also simulate the atoms within the human brain. Most simulations give at best a rough approximation of the behavior of the systems they simulate.

The same goes for simulations of the whole universe. To date, most cosmic simulations focus on the development of galaxies, typically laying a mesh over an area of the cosmos that divides it into huge units (or cells). The simulation indicates how these cells evolve and interact over time. In some systems, the size of the mesh is flexible, so that cells can become smaller in certain areas for a more fine-grained analysis. But it is rare for a cosmic simulation to descend to the level of simulating individual stars, let alone planets or organisms on those planets.

Within the next century, however, we may construct reasonably

accurate simulations of human brains and behavior. Sometime after that, we might have plausible simulations of a whole human society. Eventually we might simulate a solar system or even a universe, from the level of atoms to the level of the cosmos. In such a system, there will be bits corresponding to every entity in the universe being simulated.

Once we have fine-grained simulations of all the activity in a human brain, we'll have to take seriously the idea that the simulated brains are themselves conscious and intelligent. After all, a perfect simulation of my brain and body will behave exactly like me. Perhaps it might have its own subjective point of view. Perhaps it will experience an environment exactly like the one I experience. At this point, we're just a step away from entertaining the hypothesis that we're living in a simulation ourselves.

## Possible worlds and thought experiments

Some simulations are based on reality, while others are not. In his 1981 book *Simulacra and Simulation*, the French philosopher Jean Baudrillard distinguished four phases of simulation according to how closely they mirror reality. The first phase is *representation*, which is the “reflection of a profound reality.” The last phase is a *simulacrum*, which “has no relation to any reality whatsoever.” Baudrillard is talking about cultural symbols and not computer simulations, but a distant cousin of his distinction can be used to classify four sorts of computer simulation as well.

Some simulations (akin to Baudrillard's representations) aim to simulate a particular aspect of reality as closely as possible, the way a map represents a territory as closely as possible. A historical simulation of the Big Bang or the Second World War aims to replicate those past events closely. A scientific simulation of water boiling aims to simulate what happens when water really boils.

Some simulations aim to simulate something that *could* happen in reality. A flight simulator usually doesn't aim to simulate a flight that

has already happened, but to simulate one that could happen. A military simulation may try to simulate what could happen to the United States if there were a nuclear war.

Some simulations aim to simulate something that *could have* happened but didn't. An evolutionary simulation might simulate what would have happened if a massive asteroid impact hadn't led to the extinction of the dinosaurs. A sporting simulation might simulate what would have happened if the United States hadn't boycotted the 1980 Moscow Olympic Games.

Finally, some simulations (akin to Baudrillard's simulacra) aim to simulate worlds that bear no resemblance to reality. A scientific simulation might simulate a world without gravity. We might try to simulate a universe with seven dimensions of space and time.

As a result, simulations are not just a guide to our actual universe. They are also a guide to the vast cosmos of possible universes. Philosophers call these *possible worlds*.

In the world (that is, the universe) we live in, I became a professional philosopher. There are nearby possible worlds in which I became a professional mathematician. There are much more distant possible worlds in which I became a professional athlete. In the actual world, Hitler became leader of Germany and there was a Second World War. There are possible worlds where Hitler never took over and the Second World War never happened. In the actual world, life developed on Earth. There are possible worlds where the solar system never formed. There are even possible worlds where there was no Big Bang.

Computer simulations can help us to explore all of these possible worlds. A cosmic simulation can simulate a universe in which our own galaxy never formed. An evolutionary simulation can simulate a version of Earth in which humans never evolved. A military simulation can simulate a world in which Hitler never invaded the Soviet Union. Eventually, a personal simulation might simulate what would have happened if I had stayed in mathematics and never moved into philosophy.

Another device for exploring possible worlds is the *thought experiment*, an experiment you carry out simply by thinking. You describe a possible world (or at least part of one) and see what follows. Plato's

cave is a thought experiment. He imagines a world where prisoners can see only shadows cast on a cave wall, and asks how their lives compare to the lives of people outside the cave. Zhuangzi's butterfly is a thought experiment. Zhuangzi describes a world in which he remembers dreaming about being a butterfly, and asks how he can know he is not a butterfly that dreams he is Zhuangzi.

Thought experiments fuel science fiction. Like philosophy, science fiction explores the world as it could be. Any given science-fiction story is a thought experiment; the author conjures up a scenario and watches what follows. H. G. Wells' *The Time Machine* conjures up a world containing a time machine and then lays out the consequences. Isaac Asimov's stories in *I, Robot* conjure up a world containing intelligent robots, and Asimov then reasons about how we should interact with them.

Ursula Le Guin's classic 1969 novel *The Left Hand of Darkness* describes a possible world where humans on the planet Gethen have no fixed gender. As Le Guin puts it in her 1976 article "Is Gender Necessary?": "I eliminated gender to find out what would be left." In an introduction to the novel, she writes:

If you like you can read [this book], and a lot of other science fiction, as a thought-experiment. Let's say (says Mary Shelley) that a young doctor creates a human being in his laboratory; let's say (says Philip K. Dick) that the Allies lost the Second World War; let's say this or that is such and so, and see what happens. . . . In a story so conceived, the moral complexity proper to the modern novel need not be sacrificed, nor is there any built-in dead end; thought and intuition can move freely within bounds set only by the terms of the experiment, which may be very large indeed.

Thought experiments yield many insights. Le Guin's thought experiment gives us insight into a possibility: It tells us something about gender as it could be. Robert Nozick's thought experiment about the experience machine gives us insight into value: It helps clarify what



*Figure 6* Ursula Le Guin’s thought experiment:  
 “I eliminated gender to find out what would be left.”

is valuable to us and what isn’t. Zhuangzi’s butterfly dream gives us insight into knowledge: What can we know, and what can’t we know?

Thought experiments can stretch the boundaries of some concepts (time and intelligence) and help delimit the boundaries of others (knowledge and value). By exploring these boundaries, they teach us something about the very nature of time, or about what it is to know something.

Thought experiments can be far-fetched, but they often teach us something about reality. Le Guin says that in writing about gender she is “describing certain aspects of psychological reality in the novelist’s way, which is by inventing elaborately circumstantial lies.” Le Guin’s Gethenians may not exist, but aspects of their nature may resonate with the lived experience of many people, including some nonbinary people. Asimov’s exploration of artificial intelligence in robots can advise us about how to interact with real AI systems once they’re developed. Pla-

to's cave helps us to analyze the complex relation between appearance and reality. This is part of why thought experiments are so central in philosophy, in science, and in literature.

## Simulations in science fiction

One especially powerful thought experiment in both science fiction and philosophy is the idea of a simulated universe. What if our universe is a simulation? What follows?

James Gunn's 1955 story "The Naked Sky" was a sequel to the story about Hedonics, Inc. described in chapter 1. Both were later included in his 1961 novel *The Joy Makers*. After apparently destroying the Hedonic Council's dream machines ("In great gobs of blue, the sky began to melt"), the characters wonder whether they're still in a machine or in reality.

How could they be sure that this was reality, not another wish-fulfillment dream from the Council-mech? How could they be sure that they had really conquered it and were not just living an illusion in a watery cell? The answer was: they could never be sure.

Gunn's passage is a contender for the first explicit statement of the simulation hypothesis: the hypothesis that we're living in a computer simulation. Admittedly, computers were new at the time, and Gunn's machines are not explicitly described as computer simulations. His "sensies," in the first story, are akin to highly immersive movies, which in later stories become perfectly convincing "realies." Computer simulations play a small role in Arthur C. Clarke's 1956 novel *The City and the Stars*, but the simulation hypothesis is not entertained there.

The two ideas—computer simulation and the simulation hypothesis—may have come together for the first time in David Duncan's obscure but sophisticated 1960 short story "The Immortals." Roger Staghorn devises a computer-simulation system, Humanac, to predict the future consequences of hypothetical events. He and a colleague, Dr. Peccary,

enter the simulation and interact with people predicted to live one hundred years in the future. They have adventures and escape by the skin of their teeth. Back in the ordinary world, they turn off the simulation. The story ends:

“I can’t help wondering,” mused Staghorn, “of whose computer we’re a part right now—slight factors in the chain of causation that started God knows when and will end . . .”

“When someone pulls the switch,” said Dr. Peccary.

The deepest development of the computer simulation idea in these early years is the novel *Simulacron-3* (also known as *Counterfeit World*), published in 1964 by Daniel F. Galouye. This complex work of simulated worlds within simulated worlds was adapted by the great German director Rainer Werner Fassbinder into the German TV production *Welt am Draht* in 1973, later released with English subtitles as the film *World on a Wire*. It appears to be the debut of the simulation hypothesis in film or TV. Fassbinder’s film was later remade into the 1999 Hollywood film *The Thirteenth Floor* and is widely credited with inspiring many other films in the simulation genre.

Premiering the same year, *The Matrix*, written and directed by Lana and Lilly Wachowski, remains the best-known depiction of the simulation idea on film. The main character, Neo (in a memorable performance by Keanu Reeves) experiences an ordinary world. He goes to work, he reads books, he hangs out at parties, more or less as we do. He has a few clues that something is strange; his world has a faint green tinge, and he has a perpetual feeling of unease. Tellingly, he has been reading Baudrillard’s book *Simulacra and Simulation*. Eventually he takes the red pill and learns that he’s been living in a computer simulation all along.

*The Matrix* was partly responsible for my own entry into the simulation arena. The directors and producers of the movie had a significant interest in philosophy, and a number of philosophers were invited to write about philosophical ideas for its official website. I accepted the invitation and in 2003 published an article there called “The Matrix as

Metaphysics,” all about how the Matrix is not really an illusion. It was an early version of some of the ideas in part 3 of this book.

In “The Matrix as Metaphysics,” I introduced my own name for the simulation hypothesis. I called it the “Matrix Hypothesis” and defined it as the hypothesis that I am and always have been in a matrix. I defined a matrix as an artificially designed computer simulation of a world.

In the same year, Nick Bostrom published his important article “Are You Living in a Computer Simulation?,” which gave a statistical argument for why we should take the simulation idea seriously. (I’ll discuss the argument in chapter 5.) In another 2003 article, Bostrom introduced the name “simulation hypothesis” for the idea. This proved to be a better name than mine; the simulation idea is universal, whereas a movie is ephemeral. In this book I’m following now-standard practice in talking of the simulation hypothesis.

## The simulation hypothesis

What exactly is the simulation hypothesis? Bostrom’s version says simply, “We are living in a computer simulation.” Mine says, “We are and always have been in an artificially designed computer simulation of a world.” I think the two are consistent. My version just makes explicit a couple of things that Bostrom’s does not. First, the simulation needs to be lifelong, or at least for as long as we can remember. Being in a simulation since yesterday doesn’t count. Second, the simulation needs to have been designed by a simulator. A computer program that popped up randomly without a simulator wouldn’t count. Both of these factors are part of the simulation hypothesis as people ordinarily think of it.

What is it to be in a simulation? As I understand this notion, it’s all about interacting with the simulation. When you’re in a simulation, your sensory inputs come from the simulation, and your motor outputs affect the simulation. You’re fully immersed in the simulation through these interactions.

At the start of *The Matrix*, Neo’s biological body and brain are in