Introduction: Adventures in Technophilosophy


My views about consciousness: More precisely, my views about the hard problem of consciousness, zombies, physicalism, dualism, and panpsychism play only a minor role in this book. The main arguments about reality are equally available to materialists and dualists about consciousness. My views about the distribution of consciousness, and
especially that machines can be conscious, play a somewhat larger role.


I give some possible paths depending on your interests: If you want to follow the narrative on Descartes’s problem of the external world and my response to it, the central chapters are 1–9 and 20–24. If your main interest is virtual-reality technology, you could read chapters 1, 10–14, and 16–20. If you’re especially interested in the simulation hypothesis, you might read chapters 1–9, 14–15, 18, 20–21, and 24. If you want an
introduction to traditional problems in philosophy, perhaps I’d focus on chapters 1, 3–4, 6–8, and 14–23. It’s also worth noting that chapter 4 presupposes chapter 3, chapter 9 presupposes chapter 8 (and to some extent 6 and 7), chapter 11 presupposes chapter 10, and chapter 22 presupposes chapter 21. Parts 4–7 can be read in any order, but part 7 presupposes much of parts 2 and 3.

Chapter 1: Is this the real life?

Lead singer Freddie Mercury sings: The video for Bohemian Rhapsody depicts all four members of Queen singing the first few lines, but in fact Freddie Mercury, who wrote the song, sang all of the parts in the opening. It seems apt that in asking whether this is just fantasy, all of the voices belong to the same person.


Ancient Indian philosophers were gripped by issues of illusion and reality: For an excellent guide to issues about illusion in Indian philosophy, religion, and literature (including Narada’s transformation), see Wendy Doniger O’Flaherty, Dreams, Illusions, and Other Realities (University of Chicago Press, 1984).

James Gunn’s 1954 science fiction story: James Gunn, “The Unhappy Man” (Fantastic Universe, 1954); collected in Gunn’s The Joy Makers (Bantam, 1961).

*Life in the experience machine:* In *The Examined Life* (Simon & Schuster, 1989, 105), Nozick himself distinguishes versions of our Knowledge, Reality, and Value Questions about the experience machine: “The question of whether to plug in to this experience machine is a question of value. (It differs from two related questions: an epistemological one—Can you know you are not already plugged in?—and a metaphysical one—Don’t the machine experiences themselves constitute a real world?).”


Here and throughout, when I give PhilPapers Survey results by saying, for example, “13 percent said they would enter the experience machine,” this is shorthand for: 13 percent of respondents indicated that they accept or lean toward this view. For broader surveys beyond professional philosophers, see Dan Weijers, “Nozick’s Experience Machine Is Dead, Long Live the Experience Machine!,” *Philosophical Psychology* 27, no. 4 (2014): 513–35; Frank Hindriks and Igor Douven, “Nozick’s Experience Machine: An Empirical Study,” *Philosophical Psychology* 31 (2018): 278–98.


*Mind Question:* Why isn’t the Mind Question included on a par with the three main questions, given that this book has a section on mind in addition to sections on knowledge, reality, and value? Mainly because (1) knowledge, reality, and value correspond to the traditional divisions in philosophy (the philosophy of mind is usually regarded as part of metaphysics, the study of reality); (2) the Mind Question fragments into a number of different questions (including “What’s the relation between minds and
bodies in virtual worlds?” and “Are virtual minds real minds?”), which I address in chapters 14-16; and (3) my answer to the Mind Question isn’t as central a plank in my virtual realism as the answers to the three other questions—though Virtual minds are real minds can be regarded as a subsidiary plank.

These six further questions each correspond to an area of philosophy: There are many other areas of philosophy: for example, the philosophy of action, the philosophy of art, the philosophy of gender and race, the philosophy of mathematics, and many areas of the history of philosophy. I touch on all of these areas along the way as well, though not in as much depth as the nine areas I’ve listed.


Disciplines founded or cofounded by philosophers: Aside from Newton, I have in mind Adam Smith (economics), Auguste Comte (sociology), Gustav Fechner (psychology), Gottlob Frege (modern logic), and Richard Montague (formal semantics).

Chapter 2: What is the simulation hypothesis?


Mechanical simulation of the San Francisco Bay: For a philosophical discussion of the


*Baudrillard is talking about cultural symbols and not computer simulations:* Since Baudrillard is not primarily talking about computer simulation, the mapping from his four levels to mine shouldn’t be taken too seriously. Baudrillard’s four levels are: “It is the reflection of a profound reality,” “It masks and denatures a profound reality,” “It masks the absence of a profound reality,” and “It has no relation to any reality whatsoever: It is its own pure simulacrum.” At some points, Baudrillard counts only the fourth level as simulation.

*The vast cosmos of possible worlds:* Philosophers love to explore possible worlds. The
11th-century Islamic philosopher al-Ghazali argued that the actual world was possible before it ever existed, because God could have created it at any time. The 17th-century polymath Gottfried Wilhelm Leibniz argued that our world is the best of all possible worlds, because God has chosen it for us. The far more pessimistic 19th-century philosopher Arthur Schopenhauer argued that our world is the worst of all possible worlds, because any world worse than ours could not continue to exist. The 20th-century American philosopher Ruth Barcan Marcus explored a logical system where everything exists in every possible world.

In his book *On the Plurality of Worlds* (Malden MA: Blackwell, 1986), the American philosopher David Lewis speculated that every possible world exists. There’s a world out there where Plato’s prisoners are watching images on the cave wall. There’s a world where Hillary Clinton won the 2016 presidential election. We just happen to be in a world where Donald Trump won. No world is more real than any other; it’s just that we’re situated in one world and not another.

Versions of this “multiverse” idea—that we live in a cosmos made up of many universes—are popular in physics and cosmology. The pioneering quantum theorist Hugh Everett’s many-worlds interpretation of quantum mechanics says that our universe is constantly branching into many universes, where different histories occur. In his book *The Life of the Cosmos* (New York: Oxford University Press, 1997), theoretical physicist Lee Smolin argues that one universe may produce many new universes that evolve by a process of “cosmological natural selection.” The cosmologist Max Tegmark has speculated that every possible universe exists at least in mathematical form: *Our Mathematical Universe* (New York; Alfred A. Knopf, 2014).

James Gunn’s 1955 story: Remarkably, Gunn’s *The Joy Makers* closely anticipates two of the most important thought experiments in recent philosophy: the experience machine and the simulation hypothesis. In a preface to a later edition, he describes how he was inspired by a 1950 *Encyclopaedia Britannica* article on the psychology of feeling.

Simulations in science fiction: Related ideas can be found in fiction from the 1920s through the 1950s. The 1929 story “The Chamber of Life” (https://www.gutenberg.org/files/25862/25862-h/25862-h.htm) by Green Peyton Wertenbaker describes a fully immersive virtual reality a little like Nozick’s experience machine. In his 1935 story “Pygmalion’s Spectacles” (https://www.gutenberg.org/files/22893/22893-h/22893-h.htm), Stanley G. Weinbaum postulated spectacles that produce a multisensory experience very much like current virtual reality. In the 1940 novella *The Invention of Morel*, by the Argentine writer Adolfo Bioy Casares, a fugitive on an island encounters apparently real people, only to find they’re projections from a recording. In the early 1950’s, the science fiction stories of Ray Bradbury and Philip K. Dick depicted mechanical worlds and miniature simulated universes. Still, none of these tales involve computers per se, or the simulation hypothesis per se.

Invited to write about philosophical ideas for its official website: “The Matrix as
Metaphysics” and many other articles were solicited by Christopher Grau, a graduate student in philosophy who worked as an editor and producer for RedPill Productions, the production company for The Matrix. They were later published in Grau’s edited collection, Philosophers Explore the Matrix (Oxford University Press, 2005). At least three other edited collections of Matrix-themed philosophy have been published: William Irwin’s The Matrix and Philosophy: Welcome to the Desert of the Real (Open Court, 2002) and More Matrix and Philosophy: Revolutions and Reloaded Decoded (Open Court, 2005); and Glenn Yeffeth’s Taking the Red Pill: Science, Philosophy and Religion in The Matrix (BenBella Books, 2003).


I will use the word “sim”: The economist Robin Hanson has introduced the related term em for beings constructing by emulating a human brain. Ems and sims are distinct: an impure sim (like Neo) is a sim but not an em, and an emulated human brain in a robot body is an em but not a sim.

Philosophers revel in distinctions: In “Innocence Lost: Simulation Scenarios: Prospects and Consequences” (2002, https://philarchive.org/archive/DAIIILSv1), the British philosopher Barry Dainton makes a number of related distinctions: hard vs. soft simulations, active vs. passive simulations, original-psychology vs. replacement-psychology simulations, communal vs. individual simulations.


No universe can contain a perfect simulation of itself: Two other objections to a universe containing a simulation of itself. (For useful discussion see https://cstheory.stackexchange.com/questions/2894/can-a-computer-simulate-itself-as-part-of-a-simulated-world.)

(1) A simulation of the universe within the universe would allow us to predict the future and then falsify it: For example, if the simulation predicts that I’ll say yes at a certain time, I can read the prediction and then say no. This is the predictability paradox discussed in notes to chapter 5 (e.g. Donald MacKay, “On the logical indeterminacy of a free choice,” Mind 69:273, pp. 31-40 (1960)). It applies only to simulations that simulate the future in advance and not to concurrent or retrospective simulations.

(2) A perfect simulation of a system must be more complex than the original system. If a system has $n$ bits of complexity, a simulation of it requires $n$ bits to represent the system and more bits in overhead to run the simulation process (an operating system, for
example). So a perfect simulation of a finite universe cannot fit within that universe. Again, this objection applies only to complete simulations in finite universes.


* Imperfect simulation hypotheses:* An imperfect simulation can perhaps be seen as a perfect simulation of an imperfect unsimulated world, where the laws of physics are already messy in a way that allows red pills, communication, or approximation. For present purposes I’ll count simulations like this as imperfect all the same. Imperfect simulation hypotheses may be empirically indistinguishable from corresponding imperfect nonsimulation hypotheses. Still, the simulation versions of specific red-pill, communication, and approximation hypotheses are much better motivated than the nonsimulation versions, and specific evidence of these imperfections would reasonably be counted as evidence that we’re in a simulation.

* Tetris and Pac-Man can be regarded as simulations:* One could also perhaps regard *Tetris* and *Pac-Man* as simulations of a digital world. In this case, it’s arguably the virtual world and what it’s simulating that coincide: that is, the world is simulating itself. Something similar goes on with John Conway’s Game of Life (discussed in chapter 8), which is often called a simulation. Perhaps this is because it could be regarded as
simulating a hypothetical physical space, but more likely it’s simply simulating digital processes. In this sense, perhaps any computer program simulates itself. (These targetless simulations that don’t simulate anything other than themselves might be the most extreme case of Baudrillard’s simulacra: simulations without an underlying reality.) In any case, I’m not invoking this very loose sense of simulation for the purpose of the simulation hypothesis.

Chapter 3: Do we know things?


*If you can’t know you’re not in a simulation:* The move from *If you’re in a simulation, there’s not a spoon in front of you to If you can’t know you’re not in a simulation, you can’t know there’s a spoon in front of you* requires what philosophers call a “closure principle”: If $p$ implies $q$ (or better, if you know that $p$ implies $q$) and you can’t know not-$p$, then you can’t know not-$q$. Every now and then, someone responds to skepticism by denying this sort of closure principle and saying something like: *We can’t know we’re not in a simulation, If we’re in a simulation, there are no spoons* (and we know this), but all the same *We can know that there are spoons*. It’s not easy to make this line work well. But it’s worth being clear that we need a closure principle to turn a “no” answer to the Reality Question along with a “yes” answer to the Knowledge Question into an argument for skepticism.

Philosophers have interpreted Descartes’s celebrated slogan in many different ways: For an interpretation that denies that the *cogito* is an inference or an argument, see Jaakko Hintikka, “*Cogito ergo sum: Inference or Performance?*,” Philosophical Review 71 (1962): 3–32.

*I am conscious, therefore I am:* It’s arguable that this is what Descartes meant, since his conception of thinking coincides closely with our conception of consciousness. It seems to include imagination and sensory perception, for example.

Consciousness could be an illusion: See Keith Frankish, ed., *Illusionism as a Theory of*
Consciousness (Imprint Academic, 2017).

Chapter 4: Can we prove there is an external world?


An idea of God as a perfect being: Descartes’s argument about the perfect idea of God wasn’t original with him. In the 11th century, Saint Anselm of Canterbury put forward related “ontological” arguments for the existence of God, which we’ll discuss in chapter 7. An argument very much like Descartes’s perfect idea argument was put forward by the 16th-century Spanish scholar Francisco Suárez.


Why do we need God here?: For a modern version of idealism that uses algorithmic information theory to avoid the need for God or an external world, see Markus Müller, “Law Without Law: From Observer States to Physics via Algorithmic Information Theory,” Quantum 4 (2020): 301.

Carnap held that many philosophical problems are meaningless “pseudo-problems: Rudolf Carnap, Scheinprobleme in der Philosophie (Weltkreis, 1928); Rudolf Carnap, The Logical Structure of the World & Pseudoproblems in Philosophy, trans. Rolf A. George (Carus, 2003). For an introduction to the Vienna Circle, see David Edmonds, The
Skeptical hypotheses are meaningless: Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* (Kegan Paul, 1921). In *Language, Truth, and Logic* (Victor Gollancz, 1936), A. J. Ayer says “Consequently, anyone who condemns the sensible world as a world of mere appearance as opposed to reality, is saying something which, according to our criterion of significance, is literally nonsensical.” In “Empiricism, Semantics, and Ontology” (*Revue Internationale de Philosophie* 4 [1950]: 20–40), Carnap says that the question of “the reality of the thing world” involves a “concept cannot be meaningfully applied to the system itself.” None of the Vienna circle members explicitly discussed the simulation hypothesis, of course.


* Other replies to external-world skeptics: Online appendix.

Chapter 5: Is it likely that we’re in a simulation?

I’ll simplify by assuming that all populations have the same size: To relax the assumption that all populations have the same size, we need only weight our figures so that when counting populations (whether simulated or unsimulated), a population of one billion (for example) counts for a thousandth as much as a population of one trillion. Then the argument will go through as before. This method allows us to choose any grouping into “populations” that we like. We could even choose populations of one. Then we could use a simpler argument, whose first premise is “One in a thousand nonsim individuals will each create a million sims.” The main reason I haven’t used this formulation is to avoid the suggestion that population-creation will be an individual rather than a collective endeavor. But if things are done collectively, all we need to do is spread the credit between individuals by some method, and the argument will go through. I’ll also assume by default that populations are reasonably robust over time, so that our population counts as creating another population even if our descendants do it in a thousand years.

The conclusion that we are probably sims: The conclusion (like all other claims about probability in this chapter) can be understood as a claim about rational confidence: We should have at least 99% confidence that we’re sims. Likewise, premise 3 can be understood as saying that our conditional confidence in “We’re sims,” given that at least 99% of beings are sims, should be at least 99%. For the argument to work, strictly speaking, one needs the claim not just that premises 1 and 2 are true but that we should accept them with 100% confidence (or with high confidence, if we tweak the figures slightly). Plausibly we can’t be 100% confident in premise 1, because of sim blockers, which we’ll discuss shortly.

Math and other complications: Regarding premise 2: Let premise 1 say that at least a
fraction $k$ (between 0 and 1) of nonsim populations will create $m$ sim populations each. Then there will be at least $km$ sim populations for every nonsim population (perhaps along with some extra deeper-level sim populations created by sim populations). So in premises 2 and 3, at least $km$ in every $km+1$ beings will be simulated, and the odds will be at least $km$ to 1 in favor of our being simulated. In the argument in the text, $k = 0.1$ and $m = 1,000$, so the resulting odds are at least 100:1 in favor, which is just over a 99% chance.

The terms must be defined so that every being is either a sim or a nonsim. If a third class of being—robots, say, in a nonsimulated world—could be created, then premise 2 could be false. It could be that nonsims create many sims but even more robots, so that most beings would be robots. However, as long as robots count as either sims or nonsims, the premise is fine.

Things are more complicated if the universe is infinite. With infinite populations, proportions aren’t well-defined. As Nick Bostrom notes in the “Simulation Argument FAQ” (https://www.simulation-argument.com/faq.html), we can address this problem by defining the proportion as a limit proportion, taking the limit of proportions over increasingly large finite populations corresponding to increasing spatiotemporal areas of the universe. For example, if each such area (beyond a certain size) has more than 99% sims, then it’s not unreasonable to infer that we’re probably sims. If so, a version of the argument that uses limit proportions in (2) and (3) remains reasonable.

A residual worry about infinite populations involves a scenario in which every nonsim population that creates 1,000 sim populations also creates 2,000 nonsim populations (robot populations in the original world, say). Applying this recursively, we’ll have an infinite explosion of both sorts of populations, with nonsim populations greatly
outnumbering sim populations in the limit. In this case, premise 2 will be false. To avoid this, in the infinite case we need a stronger version of premise 1, ensuring that both sim and nonsim populations create many more sim populations than nonsim populations on average, thereby ensuring that sims greatly outnumber nonsims in the limit.

*000 Sim blockers: Deleted.


An objection related to Intelligent sims are impossible is Conscious sims are impossible. The way I’ve defined sims here, sims have to be intelligent, but they don’t have to be conscious, so this objection doesn’t threaten premise 1. If nonsims create conscious sims, that’s good enough for the premise to be true. Instead, we’ll consider this worry shortly, under objections to premise 3.

*000 On current estimate, the brain’s computing speed is around 10 petaflops: One rough calculation assumes 100 billion neurons with around 1,000 connections (or synapses) each, where each synapse can be encoded with 10 bits of information each. This adds up to about 100 terabytes of information. If each synapse transmits a signal up to 100 times a second, and we equate a single synaptic transmission with a single floating-point operation (or flop) in a computer, the brain as a whole will perform about $10^{16}$ flops per second.


*Computronium:* The name “computronium” was introduced for the idea of programmable matter by Tommaso Toffoli and Norman Margolus; see their “Programmable matter: Concepts and realization,” *Physica D*, 47, no. 1–2 (1991): 263–72; and Ivan Amato, “Speculating in Precious Computronium,” *Science* 253, no. 5022 (1991): 856–57. The now-common usage for maximally efficient programmed matter was popularized in science-fiction works such as Charles Stross’s *Accelerando* (Penguin Random House, Ace reprint, 2006), in which much of the solar system is turned into computronium.

*If we’re in a simulation, evidence about our computer power may be misleading:* For versions of this objection, see Fabien Besnard, “Refutations of the Simulation
Argument,” http://fabien.besnard.pagesperso-orange.fr/pdfrefut.pdf, 2004; and Jonathan Birch, “On the ‘Simulation Argument’ and Selective Scepticism,” Erkenntnis 78 (2013): 95–107. At worst, we can reason: (1) either our evidence about computer power is heavily misleading, or it is not, (2) if our evidence about computer power is heavily misleading, we’re probably in a simulation (as that’s the most likely way for this evidence to be misleading), (3) if our evidence about computer power is not heavily misleading, we’re probably in a simulation (by the original argument), so (4) we’re probably in a simulation. Still, the likelihood that simulations will be misleading does bring out that the simulation argument can easily be turned into an argument for skepticism about certain sorts of scientific knowledge, even if (as I will argue) it doesn’t lead to global skepticism about the external world.


Nonsims will die before creating sims: Another quite different way that this could be true is if the vast majority of human-level nonsims are Boltzmann brains (discussed in chapter 24), all of which will almost certainly dissolve within seconds.

A distinctive version of the Nonsims will die before creating sims sim blocker is We are alone. This blocker arises if we’re the only nonsim population in the cosmos and we die before we’re able to create sims. (There’s a parallel We are alone version of We’ll choose not to create sims to which similar issues apply.) In this case, premise 1 will be false.

Zero percent of nonsims will create sims. This won’t require strong sim blockers that make it near-inevitable that intelligent populations will die before producing simulations. It suffices that there’s a 50-50 chance (say) that intelligent populations die off and there’s
a single nonsim population that turns out to be unlucky. It’s perhaps harder to exclude this version of *We’ll all die first* than the version that applies to multiple populations, depending on one’s confidence that there will be a single nonsim population in a universe as large as ours, along with one’s confidence that such a population won’t create sims. If we regard these hypotheses as likely, this will significantly reduce the probability that we’re sims. If we regard them as unlikely, they won’t affect the probability by much. There’s also a version of the *We are alone* objection that combines with the *We know we’re not the sims we create* objection, so that even if we create sims, we can’t go from *Most beings are sims* to *We’re probably sims*, since we created the sims in question and therefore we know we’re not them.

*000  *Simulate the decision first and see how things go:* The practical role for simulations in decision-making may be subject to some limits, as I discuss in chapter 7. Simulations for decision-making also may not lead to sims with our experiences. The beings using simulation technology presumably live in an era more advanced than ours, and it’s not obvious how simulating more primitive beings will be useful for their decision-making.

*000  *We could be nanoscale nonsims:* The trouble with this objection is that the nanoscale physical environment is very far from being a shrunken version of the ordinary physical environment, since the nanoscale environment is dominated by molecules and quantum effects. This might be fine for some purposes, such as using nanoscale robots to perform nanoengineering or nanoscale brains to solve mathematical problems. Perhaps it could even turn out that nanoscale nonsims of this sort will outnumber sims. But humanlike sims will still outnumber humanlike nonsims, which is what matters. (For a more optimistic take on miniature universes, see the discussion of “type-3 simulations” in

A loophole is that we’re assuming a universe with physical laws like ours. Perhaps there’s a somewhat different world, where humanlike nonsims create nanoscale humanlike robot nonsims, which are as cheap and easy to make as sims and outnumber them. Perhaps we could even be those nonsims. So we can’t entirely rule out the possibility that we’re nanoscale nonsims created in a world where these are cheap and easy to create. Something similar applies to a world with infinite space or baby universes where nonsims are cheap and easy to create.

*Interestingness is a sim sign:* Robin Hanson, “How to Live in a Simulation,” *Journal of Evolution and Technology* 7 (2001).

*Our position early in the universe is a sim sign:* Carl Shulman has suggested to me that this sim sign is responsible for the distinctive strength of the ancestor simulation argument. There are specific reasons to create early-universe simulations (e.g., interest in one’s history), and it’s relatively easy to do so. Also, the population of the early universe is relatively small and it’s easy for it to be outnumbered by sims. These points do not generalize to the later universe, which may have a far huger population and will be much harder to simulate.

*Sim sign:* In his “The PNP Hypothesis and a New Theory of Free Will” (*Scientia Salon*, 2015), Marcus Arvan argues that a version of the simulation hypothesis is the best explanation of free will and of various features of quantum mechanics, suggesting in effect that these phenomena are sim signs.

*I discuss five more objections related to nonsim signs:* Online appendix.
Simulators will avoid creating conscious sims: Thanks to Barry Dainton, Grace Helton, and Brad Saad for versions of this suggestion. In her “Epistemological Solipsism as a Route to External World Skepticism” (Philosophical Perspectives, forthcoming), Helton argues that ethical simulators may well create simulations in which only one being is conscious—in which case, any conscious being should take seriously the solipsistic thesis that they are the only conscious being in the universe.

Sims won’t experience large universes: In addition to suggesting that complex physics is a nonsim sign, the physicist Frank Wilczek (“Are We Living in a Simulated World?”, Wall Street Journal, January 9, 2020) has suggested that the continuous physics of our world is a nonsim sign: it’s inefficient to simulate continuity digitally, so most sims will live in worlds with digital physics. However: Once we acknowledge the possibility of analog simulations, it’s not clear that most sims will live in digital worlds and it’s not clear that continuity is a sim sign.

Simulation that takes shortcuts: Online appendix. To what extent can simplified models be used to simulate the behavior of macroscopic objects in a way consistent with all of our observations? To handle every possible observation of a system, simplified models won’t be enough; a simulation in full detail will be required. But most actual systems are observed less closely than this. For example, if a bowl of ice slowly melts into a bowl of water with no-one watching for a day or so, a simple model specifying the water temperature a day later may suffice. If someone is watching as the ice melts, a more detailed model of the melting process will be required. If images are recorded for possible later examination and scientific analysis, a far more detailed model will be required.
Simulators seeking efficiency in modeling worlds like ours will presumably use models at different levels, depending on the level of observation involved. But if a system leaves many observable traces on systems around it (which may be the typical case), and those traces can be analyzed, similar issues will arise. It will be risky to use a simplified model to simulate a hurricane, for reasons like this. The effects of the simplified model will differ in subtle ways from the effects of a genuine complex hurricane, and these effects will in principle be analyzable in a way that could give away shortcuts in the simulation. If simulators have control over what sort of observations are made when, then this will give them much more leeway to use simplified models.

Julian Togelius has suggested to me that for related reasons, quantum mechanics may be a sim sign. There are versions of quantum mechanics suggest that reality only becomes determinate when we are conscious of it (see e.g. chapter 14). This is what one would expect in a just-in-time simulation where simulators only simulate what is necessary to explain sim’s conscious observations. On the other hand, simulating an uncollapsed quantum wave-function may not be any easier than simulating a collapsed version.

*000  Major sim signs: Requiring humanlike sims to have exactly the same precise sim/nonsim signs as humans might mean that there are no other humanlike beings. For our purposes, we can individuate sim signs as broadly as possible (e.g., experiences a large universe, vs. experiences a specific large universe), when the details make little difference to the probabilities. These major stamps are potentially widely shared in other populations.

*000  I don’t think Bostrom’s formula or his conclusions are quite correct as they stand: Online appendix: Bostrom on the simulation argument.

*000  If there are no sim blockers, we are probably sims: The conclusion can be read as saying
that conditional on there being no sim blockers, we should be confident that we are sims (that is, that we should have a high conditional probability in *We are sims*, given *There are no sim blockers*). Premise 2 should be read with the same structure. Premise 1 works best if understood as the claim that conditional on there being no sim blockers, we should be certain that we are sims.

*000* Both premises now require only relatively small assumptions: Premise 1 requires only the plausible assumption that if nothing prevents the creation of many humanlike sims (enough of them that most humanlike beings are sims), then there will be many humanlike sims. Premise 2 requires only a version of the indifference principle, discussed earlier.

One might worry that the argument is now almost trivial, in that an argument like this will be valid even if “sim” is replaced by anything else. For example, we could argue that if there are no redhead blockers, then most humanlike beings are redheads, so we are probably redheads. The difference is that where sims are concerned, it looks like there’s a strong incentive to create many humanlike sims, as well as cheap and easy means to do so. Given this, humanlike sims will be created unless something prevents this from happening. That’s a version of the assumption required for premise 1. It’s not clear that there’s the same strong incentive to create redheads, in which case the corresponding assumption and premise may be false: It’s not the case that redheads will be created unless something prevents them from being created.

There may be other cases with incentives and means so that the assumption is true. One could argue that if there are no robot blockers, we are probably robots. In this case, there’s arguably incentive and means, so that versions of both premise 1 and 2 will
be true. But in this case there’s arguably an obvious robot blocker; namely, creating sims for the same purposes will be much cheaper and easier. What’s distinctive about the sim version of the argument is that there’s no really obvious sim blocker, so that the existence of sim blockers would be more surprising and interesting.

*000 If it came up heads, he connected me to a perfect simulation. We could also adapt a classic thought experiment by the philosopher Carl Ginet. We’re driving down the road, and we see what appears to be a barn. However, we’re told that in one-quarter of the counties in the area, selected randomly, all barns have been replaced by fake barns, which have just the façade of a barn with nothing behind it. From the road, the fake barns are indistinguishable from real barns. If so, we certainly cannot know that the barn we’re seeing is a real barn, even if it is in fact a real barn. Once we know there’s it is a serious possibility that we’re in fake-barn country, no philosophical maneuvers can remove this possibility and allow us to know that we are seeing a barn.

<NH1>Chapter 6: What is reality?

<NTX>*000 Virtual Realism: Writing at the same time as Heim, Philip Zhai also argues for a sort of virtual realism in his 1998 book Get Real: A Philosophical Adventure in Virtual Reality. I discuss Heim’s and Zhai’s versions of virtual realism at more length in an online appendix. Other authors whose work contains elements of virtual realism include David Deutsch (discussed in chapter 6) and Philip Brey (discussed in chapter 10). Elements of simulation realism are endorsed by Douglas Hofstadter (discussed in chapter 20) as well as in the articles by Andy Clark and Hubert Dreyfus in Philosophers Explore the Matrix. In addition, O. K. Bouwsma (chapter 6) and Hilary Putnam (chapter 20) show sympathy for a view akin to simulation realism without explicitly discussing
simulations per se.


Eleatic dictum: This dictum was named the “Eleatic principle” by Graham Oddie, in “Armstrong on the eleatic principle and abstract entities,” Philosophical Studies, 41:2, pp. 285-95 (1982), following a discussion by David Armstrong (1978). Jaegwon Kim (1993) calls a closely related principle (“To be is to have causal powers”) “Alexander’s dictum”, after the Australian philosopher Samuel Alexander, but it’s hard to find the principle explicitly in Alexander. The British philosopher L. Susan Stebbing put forward a version of the principle in “The Philosophical Importance of the Verb ‘To Be,’” Proceedings of the Aristotelian Society, 18, pp. 582-89 (1917-1918): “The real is to be defined in terms of causal efficacy.”

Austin’s lectures: J. L. Austin, Sense and Sensibilia (Oxford University Press, 1962).

There are other strands we could have added: Other strands include Reality as observability. Reality as measurability. Reality as theoretical utility. (These are related to the causal power strand.) Reality as authenticity. Reality as naturalness. Reality as originality. Reality as fundamentality. (These are related to the genuineness strand.) Then there are the senses of “really”—what we mean when we say that something is really the case. Here the strands include Reality as truth. Reality as actuality. Reality as factuality. (These are related to the non-illusoriness strand.) Reality as objectivity. Reality as intersubjectivity. Reality as evidence-independence. (These are related to the mind-independence strand.) Each of these senses of “really” arguably yields a corresponding
sense of “real” by translating “X is real” into “X really exists.” (I set aside the strands at play in “real number” and “real estate”—though it’s worth noting that the terminology of real and imaginary numbers comes from Descartes!) Of these many strands, perhaps those that most threaten the status of simulated objects as real are some of those in the genuineness strand, such as Reality as originality and Reality as fundamentality, which I discuss in the text. For further discussions of the many senses of “real,” “really,” and “reality,” see Jonathan Bennett, “Real,” *Mind* 75 (1966): 501–15; and Steven L. Reynolds, “Realism and the Meaning of ‘Real,’” *Noûs* 40 (2006): 468–94.

*000 Striking how uncommon this view has been: It would be natural to find versions of the no-illusion view in various idealists, phenomenalists, and pragmatists. There are occasional passages that make general claims about appearance and reality that seem to indirectly commit them to something like the view (e.g., David Barnett pointed me to passages in Arthur Schopenhauer’s *The World as Will and Representation*, and Griffin Klemick to passages in C. I. Lewis’s *Mind and the World Order*), but it’s surprisingly hard to find explicit statements of this view about skeptical scenarios. I’m interested to hear of sources.


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

First interesting argument for the existence of god in a long time: https://www.simulation-argument.com/.

The fine-tuning argument is controversial: In a 2020 PhilPapers Survey question about what explains fine-tuning, 17 percent said design explains it, 15 percent said a multiverse
explains it, 32 percent said it’s a brute fact, and 22 percent said there’s no fine-tuning.

*Naturalism:* There are many varieties of naturalism. Metaphysical naturalism carries the minimal commitment that everything is part of nature. Sometimes naturalism is understood as a version of materialism, which holds that everything is physical; but there are many people who reject materialism and accept naturalism, including me.

Methodological naturalism is sometimes understood as giving a central role to science in doing philosophy and in understanding the world. Leading 20th-century naturalists include W. V. Quine, Rudolf Carnap, and more recently figures such as Patricia Churchland and Daniel Dennett.


Simulation afterlife: For an optimistic perspective, see Eric Steinhart’s Your Digital Afterlives: Computational Theories of Life after Death (Palgrave Macmillan, 2014).


Chapter 8: Is the universe made of information?

Leibniz invented the bit: Gottfried Wilhelm Leibniz, “De Progressione Dyadica” (manuscript, March 15, 1679); “Explication de l’arithmétique binaire,” Memoires de l’Academie Royale des Sciences (1703). It is sometimes said that the I Ching inspired Leibniz’s discovery. In fact, he formulated binary arithmetic some years before Joachim Bouvet introduced him to the I Ching and pointed out the resemblance, after which Leibniz built it into his exposition. There is also a case for Thomas Hariot inventing the bit a century before Leibniz: See John W. Shirley, “Binary Numeration before Leibniz” (American Journal of Physics 19, no. 8 [1951]: 452–54). The 20th-century American mathematician Claude Shannon, who cointroduced the label “bit,” is sometimes called the “inventor of the bit.” As we’ll see, what Shannon invented was an information-theoretic measure and not the binary digit.

You can try out the Game of Life: playgameoflife.com. The default starting point is a glider, but you can try many other arrangements, including a glider gun: playgameoflife.com/lexicon/Gosper_glider_gun.

Metaphysical theorizing: For these historical metaphysical systems, see A. Pablo Iannone, *Dictionary of World Philosophy* (Routledge, 2001).

An oscillation among materialism, dualism, and idealism: In the 2020 PhilPapers Survey, 52 percent accept physicalism about the mind while 22 percent reject it. In a question about consciousness, 22 percent accept dualism and 8 percent accept panpsychism (33 percent accept functionalism, 13 percent accept the mind-brain identity theory, and 5 percent accept eliminativism, which we haven’t discussed here). In a question about the external world, 7 percent accept idealism (5 percent accept skepticism and 80 percent accept non-skeptical realism).


Structural, semantic, and symbolic information: See an online appendix for more in-depth discussion. This is my own way of dividing up the territory, but related distinctions have been made many times before. There are many different taxonomies of information: See, for example, Mark Burgin, *Theory of Information: Fundamentality, Diversification and Unification* (World Scientific, 2010); Luciano Floridi, *The Philosophy of Information*
*000 *Shannon information:* Somewhat confusingly, Shannon used the word “bit” for his measure of structural information as well as for binary digits. In his 1948 “A Mathematical Theory of Communication” [The Bell System Technical Journal, 27, pp. 379-423, 623-656, 1948; reprinted as *The Mathematical Theory of Communication* (Champaign IL: University of Illinois Press, 1949)], he says: “The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits*, a word suggested by J. W. Tukey.”

Shannon’s definition enshrines an essential ambiguity: “Bit” can be used either for a digit or a measure. It seems that Shannon’s Bell Labs colleague John W. Tukey introduced the term *bit*, explicitly for binary digits, in a 1947 memorandum. Shannon’s 1948 article extends it to a measure, while also saying that bits are digits: See Henry S. Tropp, “Origin of the Term *Bit*,” *IEEE Annals of the History of Computing*, 6:2, pp. 152-55 (1984). There are further ambiguities in both of these notions. As we’ll see, even setting measures aside, “bit” is ambiguous, meaning either a purely mathematical entity (a *binary number*, 0 or 1) or a physically embodied entity (a *binary state*: that is, a physical state with either of two values, labeled 0 and 1). “Bit” as a measure is also ambiguous, meaning either the number of binary digits in a structure (often used as a unit of storage, as in a 256GB memory), or the amount of Shannon information in a structure. Sometimes the latter measure is called a *shannon* rather than a bit, for clarity. For my purposes, most relevant is the use of “bit” for a physically embodied entity (a binary state, which is
physically embodied structural information).


Continuous digits: The terms “continuous-valued digit” and “analog digit” are sometimes used in the literature (e.g., Saed et al., “Arithmetic Circuits for Analog Digits”), but as far as I know, there’s no standard abbreviation. Ant and cont are unlovely terms, so I’m reluctantly using real despite its imperfect connotations. For example, while it suggests purely mathematical real numbers, physically realized reals (as with bits) are more crucial for our purposes. (Also, reals shouldn’t be confused with real in the sense of reality; and continuous quantities are often complex numbers rather than real numbers.) Whereas a bit is physically embodied as a binary state in a physical system, a real is physically embodied as a real-valued state in a physical system (where both are individuated in a substrate-neutral way). Note that there isn’t really a measure of the quantity of continuous information analogous to Shannon-style bit measures, in part because multiple reals can be recoded as a single real and vice versa.

*000 Structural information can be physically embodied: Structural information can also be
mentally embodied: For example, a state of consciousness with patterns of light and dark embodies a structure of bits. In *The Conscious Mind*, I argued for a double-aspect theory of (structural) information, where the same structural information is simultaneously embodied physically and mentally.

*A difference that makes a difference:* Gregory Bateson, *Steps to an Ecology of Mind* (Chandler, 1972). Bateson gives credit to Donald Mackay, who said, “Information is a distinction that makes a difference.”

*Physical information:* This discussion brings out that structural information comes in *abstract* (mathematical) and *concrete* (physical and mental) varieties. The same is arguably true of semantic information. Facts and propositions are abstract semantic information. But facts and propositions can be *believed, asserted, known, written, encoded*, and so on. We can think of these beliefs, assertions, etc., as concrete (physically or mentally embodied) semantic information. In this sense, symbolic information (bits encoding facts and propositions) is one variety of concrete semantic information, at least if concrete bits do the encoding.

*Information is physical:* this slogan was put forward by the physicist Rolf Landauer in “Information Is Physical,” *Physics Today* 44, no. 5 (1991): 23–29.


*Wheeler’s powerful slogan “it from bit”:* John Archibald Wheeler, “Information, Physics, Quantum: The Search for Links,” *Proceedings of the 3rd International Symposium on the*
Space and time emerge from something more fundamental: I discuss this idea in “Finding Space in a Nonspatial World,” in Philosophy beyond Spacetime, eds. Christian Wüthrich, Baptiste Le Bihan, and Nick Huggett, which contains many other discussions of emergent spacetime (Oxford University Press, 2021).


It-from-bit-from-it: For related discussion, see Anthony Aguirre, Brendan Foster, and Zeeya Merali, eds., It from Bit or Bit from It? On Physics and Information (Springer, 2015); and Paul Davies and Niels Henrik Gregersen, Information and the Nature of Reality (Cambridge University Press, 2010).


Reality grounded in continuous information: What’s the difference between the pure it-
from-real thesis and the standard thesis that physics involves continuous quantities? The extra content comes from the structuralist thesis requiring that the continuous physics in it-from-real theories be put in substrate-neutral mathematical form. Standard Newtonian physics is not yet in this form. It makes essential appeal to mass, distance, and so on. Mass and distance can be represented as continuous values, but they aren’t substrate-neutral, whereas reals are. It-from-real Newtonian physics requires putting the theory in mathematical form without requiring any special substrate mentioning mass and distance. It-from-real-from-it physics allows these reals to be further grounded in specific substrates, possibly including substrates specific to mass or distance. It’s plausible that any continuous physical theory can be represented as an it-from-real-from-it theory in this fashion (though complex values rather than real values are often required). Pure it-from-real physics (like pure it-from-bit physics) holds, more radically, that these pure reals are fundamental. It’s the pure it-from-real thesis that corresponds to a sort of ontological structural realism.

Chapter 9: Did simulation create its from bits?

I need only establish that the simulation hypothesis leads to the it-from-bit creation hypothesis: The reverse claim—“If the it-from-bit creation hypothesis is true, the simulation hypothesis is true”—raises a number of further issues (what about the pure it-from-bit hypothesis? what about the computer?), some of which I’ll discuss in a subsequent note. The general moral is that one has to understand computer simulations and the simulation hypothesis in an inclusive way for the equivalence to hold in both directions.

Simulation run on a quantum computer: For discussions of simulated worlds in the

*000* The bits our simulator is creating aren’t fundamental: This point raises more of an objection to the reverse claim that the it-from-bit creation hypothesis leads to the simulation hypothesis. Is the pure it-from-bit creation hypothesis consistent with the simulation hypothesis? Doesn’t the simulation require a programmable computer, where bits are nonfundamental? To accommodate these versions of the it-from-bit hypothesis, we need to understand computer simulations so that they require only a system executing the relevant algorithm, whether or not the system involves nonfundamental bits (connecting up the bits in a pure it-from-bit world still counts as a simulation), and whether or not it involves a programmable computer (which may be hard-wired for one algorithm only). Arguably, if our experiences come from algorithmic systems of this sort, this should still count as a version of the simulation hypothesis. If someone resists this expansive understanding of the simulation hypothesis, we can qualify the reverse claim so that only certain versions of the it-from-bit creation hypothesis lead to the simulation hypothesis. The main argument works either way.

*NH1* Chapter 10: Do virtual reality headsets create reality?


000 Attempts at a Metaverse: The leading platforms for social VR at the time of writing in early 2021 included VRChat, Rec Room, Altspace VR, Bigscreen, and Facebook Horizon. *Second Life* has proved difficult to transport from 2-D computer screens to VR because the required frame rate is too high.

Charles Sanders Peirce enshrined this definition: C. S. Peirce, “Virtual,” in *Dictionary of Philosophy and Psychology*, ed. James Mark Baldwin (Macmillan, 1902). Peirce goes on to say that this meaning of “virtual” as *in effect* should be distinguished from the sense where it means *potential*, as with an embryo that is a potential person. An embryo does not have the power of a person, so it is not a virtual person in the “in-effect” sense, but it has the power to become a person, so it is a virtual person in the *potential* sense. The idea of virtuality as potentiality is no longer central in ordinary uses of the word, but it has led to an important philosophical tradition associated with the French philosophers Henri Bergson (in his 1896 book *Matter and Memory*) and Gilles Deleuze (in his 1966 book *Bergsonism* and other works). As Deleuze puts it: “virtual” (in his sense) is opposed not to “real” but to “actual,” where “actual” is understood in the sense of *actualization*. The virtual is not yet actualized (like an embryo), or is in the process of being actualized (like a crystallization), or was once actualized (like a memory). For a guide to the many senses of virtuality, see Rob Shields, *The Virtual* (Routledge, 2002).


* Early uses of “virtual reality” and “virtual world”: Online appendix.

We call it a merely virtual object: Susanne K. Langer, *Feeling and Form: A Theory of Art* (Charles Scribner’s Sons, 1953), 49.


Made of atoms: Philosophers have understood the sense in which physical objects are


When is a virtual X a real X?: More precisely a virtual X is a real X as long as X is a causal/mental invariant: something that depends only on the abstract causal organization and the mental properties of a situation (see “The Matrix as Metaphysics” and “The Virtual and the Real”, op. cit.). Philip Brey (“The Social Ontology of Virtual
Environments,” (op. cit.), “The Physical and Social Reality of Virtual Worlds,” in Mark Grimshaw, ed., *The Oxford Handbook of Virtuality* (Oxford University Press, 2014)) addresses the same question and answers that a virtual X is an X if and only if X is an institutional kind (such as money), one that is constituted by collective social agreements in the right way. I think that the “only if” claim is not quite right: virtual calculators are calculators and virtual boredom is boredom, where both are causal/mental kinds though neither are institutional kinds. But it is plausible that most institutional kinds are causal/mental kinds, so Brey’s “if” claim is plausible.

<NH1>Chapter 11: Are virtual reality devices illusion machines?


000 Plausibility Illusion: A philosopher might have called this the Event Illusion or the Happening Illusion because it centers on the sense that certain events are really happening.

VR is not an illusion: Philip Zhai also argues against the Illusion View of VR in his 1998 book Get Real: A Philosophical Adventure in Virtual Reality (Rowman & Littlefield), discussed in an online appendix.


**Experiencing a real virtual body:** Research on virtual worlds suggests that users often
(but not always) like to adopt avatars that express the bodies of their “ideal selves”—roughly, the bodies they would like to have or that they identify with. It’s common for people to adopt bodies that they take to be more attractive than their own.

In the physical world, our bodies deeply affect the way we behave. The same is true in virtual worlds. People tend to behave in a way that fits their avatar. VR researchers Nick Yee and Jeremy Bailenson call this the Proteus effect, named after the shape-shifting Greek God Proteus. For example, people with avatars that they perceive as more attractive are more likely to walk closer to others and to engage in more self-disclosure. People with taller avatars are more likely to behave self-confidently. [Yee & Bailenson, “The Proteus Effect: The Effect of Transformed Self-Representation on Behavior,” *Human Communication Research*, 33, pp. 271-90 (2007).] There’s some evidence that experience with an avatar can affect nonvirtual behavior as well. For example, white people, given a black avatar, show (in some circumstances) a reduction in implicit racial bias that can persist for at least a week afterwards. [Domna Banakou, *et al.*, “Virtual Embodiment of White People in a Black Virtual Body Leads to a Sustained Reduction in Their Implicit Racial Bias,” *Frontiers in Human Neuroscience*, (2016) https://doi.org/10.3389/fnhum.2016.00601.] All this seems consistent both with the illusion and no-illusion views of body ownership.

Chapter 12: Does augmented reality lead to alternative facts?


Relativism is a deeply controversial idea: Maria Baghramian, *Relativism* (Routledge,
Chapter 13: Can we avoid being deceived by deepfakes?

Henry Shevlin posted an interview online: Public version is available at https://www.facebook.com/howard.wiseman.9/posts/4489589021058960. Thanks to Henry Shevlin for permission to use this.

Deepfakes can be found in contexts: Sally Adee, “What Are Deepfakes and How Are They Created?,” IEEE Spectrum (April 29, 2020).


The term “fake news” has become controversial: See Josh Habgood-Coote, “Stop Talking about Fake News!,” Inquiry 62, no. 9–10 (2019): 1033–65; and Jessica Pepp,
Eliot Michaelson, and Rachel Sterken, “Why We Should Keep Talking about Fake

Fake news isn’t the same as false news: On the definition of fake news, see Axel Gelfert,
Fake News?,” *Versus* 2, no. 127 (2018): 207–27; and Don Fallis and Kay Mathiesen,

Interconnected in webs of mutual endorsement: For a network analysis of fake news and
other misinformation, see Cailin O’Connor and James Owen Weatherall, *The
Misinformation Age: How False Beliefs Spread* (Yale University Press, 2019).


<NH1>Chapter 14: How do mind and body interact in a virtual world?

<NTX>000 Second-ever conference on artificial life: Christopher G. Langton, Charles Taylor,
J. Doyne Farmer, and Steen Rasmussen, eds., *Artificial Life II* (Santa Fe Institute, 1993).

000 Alan Kay’s Vivarium: Larry Yaeger, The Vivarium Program,”

000 It struck me that these creatures would almost certainly become dualists about the mind:
David J. Chalmers, “How Cartesian Dualism Might Have Been True,” February 1990,
https://philpapers.org/rec/CHAHCD.

000 Dualism can be found in many different cultures. Kwame Gyekye, “The Akan Concept of
a Person,” *International Philosophical Quarterly* 18 (1978): 277–87, reprinted in

*Descartes articulated a classic form of dualism:* René Descartes, *Meditations on First Philosophy* (Meditations 2 and 6, 1641) and *Passions of the Soul* (1649), both excerpted in *Philosophy of Mind*, ed. Chalmers.


*Biological and virtual brains synchronized:* This is a little reminiscent of Leibniz’s theory where there is preestablished harmony between mind and body, although Leibniz’s picture avoided any causal interaction between the two.


*<NH1>Chapter 15: Can there be consciousness in a digital world?*

*<NTX>000 Mind uploading:* Russell Blackford and Damien Broderick, eds., *Intelligence*


Other minds: In the 2020 PhilPapers Survey, 89 percent of respondents say cats are conscious, 35 percent say flies are conscious, 84 percent say newborn babies are conscious, 3 percent say current AI systems are conscious, and 39 percent say future AI systems can be conscious (while 27 percent deny this, and the rest adopt various forms of neutrality).


We might become the machine: In general, cases of this sort provide some of our best evidence about the problem of other minds. Suppose someone suggests that as a matter of principle, no one in New York City is conscious. I could try to observe people in New York City from the outside to tell whether or not they’re zombies, but there will be obvious limitations. Alternatively, I could simply go to New York City, and if I find myself conscious there I have falsified their hypothesis: People in New York City can be conscious! Of course, this may not be such conclusive evidence for people outside the city watching me. And after I leave the city, I may find myself wondering how I know that my memories of being conscious in the city are accurate. Still, most of us would find this sort of first-person evidence to be among the most convincing evidence we can have for any hypothesis about consciousness.

Many people accept that the original person dies: In a 2020 PhilPapers Survey question on “Mind uploading (brain replaced by digital emulation),” 27 percent of professional philosophers said this is a form of survival and 54 percent said it’s a form of death.

Chapter 16: Does augmented reality extend the mind?


The Extended Phenotype: Richard Dawkins, The Extended Phenotype (Oxford University Press, 1982).


Report AFOSR-3233, Stanford Research Institute, October 1962.


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

*000 Nozick’s 1974 fable of the experience machine: Aside from James Gunn’s story “The Unhappy Man,” described in chapter 1, there are many other antecedents for the experience machine in fiction: E. M. Forster’s 1909 short story “The Machine Stops” (the machine is not fully immersive, so the story arguably anticipates the internet and videoconferencing more than it anticipates the experience machine), Green Peyton Wertenbaker’s 1929 story “The Chamber of Life,” the “dream machines” in Laurence Manning and Fletcher Pratt’s 1930 story “City of the Living Dead”, the “feelies” in Aldous Huxley’s 1932 novel Brave New World, Arthur C. Clarke’s 1949 novella “The Lion of Comarre,” John MacDonald’s 1950 story “Spectator Sport” (whose company, World Senseways, is reminiscent of Gunn’s Hedonics, Inc.), and more. In 1968 there was almost a Star Trek episode based on a scenario like the experience machine: The Joy Machine, written by Theodore Sturgeon, was to be the third season’s 25th episode, but the show was cancelled after the 24th episode. Later, James Gunn himself adapted the episode as a novel.
*VR is not preprogrammed:* As Dainton puts things in “Innocence Lost”:

“The virtual lives sustained by experience machines are of the *passive* kind: they consist of solitary streams of consciousness that are completely controlled and preprogrammed. . . [N]ot all virtual lives need be like this. Of particular interest here are AC-simulations, i.e., virtual lives that are both *active* and *communal*, in the senses introduced above. Subjects in AC-simulations possess their own autonomous psychologies (whether original or replacement). They lead their own lives: their actions are not pre-programmed (they are as free as anyone can be). And they can causally interact with other subjects in their virtual environment (and these other subjects are autonomous individuals in their own right, rather than merely the appearances of such). Given all this, it is hard to see why life in an AC-simulation should be regarded as being inherently less valuable or worthwhile than a normal life.”

Similarly, Cogburn and Silcox, in “Against Brain-in-a-Vatism,” write:

“What Nozick is describing might be called “passive virtual reality.” . . . [I]n active, non-
solipsistic VR, the player is not only “doing something,” but is also “being” a certain way, at least insofar as her/his behavior has ethically significant effects upon the lives of others.”

Cogburn and Silcox reject the “brain-in-a-vatism” attitude toward VR, which holds that life in VR has only the limited value and epistemic status of the life of a brain in a vat. By my lights, they’re right about VR, at least where the Value Question is concerned (on the Reality Question, they defend a form of fictionalism, albeit one that allows us to learn truths from fiction), but they are too pessimistic about brains in vats.


1863 book: John Stuart Mill, Utilitarianism (Parker, Son, & Bourn, 1863).


<NH1>Chapter 18: Do simulated lives matter?


Euthyphro’s dilemma: This is one case where the Simulation Riposte makes the argument all the stronger. We can imagine creating a world with simulated counterparts of Socrates.
and Euthyphro. Sim Socrates asks Sim Euthyphro: Is this the right thing to do because the
 gods command it, or do the gods command it because it is the right thing to do? We’re
 the gods of the simulation, so perhaps we can answer the question. If we’re non-
 interfering creators, we might answer “neither”—we haven’t commanded any actions at
 all. If we’re activist creators who lay down laws for the simulation, then perhaps our
 commands make certain actions legal or illegal in the simulation. But our commands
don’t make things right and wrong. If we commanded torture, it wouldn’t be right. And if
 we want to make people in the simulation do the right thing, we have to figure out the
 right thing to do.

Classic 1958 article: G. E. M. Anscombe, “Modern Moral Philosophy,” Philosophy 33,

Virtue ethics has recently had a resurgence: In the 2020 PhilPapers Survey, 32 percent of
respondents endorsed deontology, 31 percent endorsed consequentialism, and 37 percent
endorsed virtue ethics. These figures are up from 26 percent, 24 percent, and 18 percent,
respectively, in 2009. The survey allowed multiple endorsements in 2020 but not 2009; in
any case, virtue ethics has moved from last to first.

A being has moral status: For a review of general issues about moral status, see
Encyclopedia of Philosophy (Spring 2021), https://plato.stanford.edu/entries/grounds-
moral-status/. On issues about the moral status of AI systems, see Matthew Liao, “The
Moral Status and Rights of Artificial Intelligence,” in The Ethics of Artificial Intelligence,
ed. Matthew Liao (Oxford University Press, 2020) and Eric Schwitzgebel and Mara
Garza, “Designing AI with Rights, Consciousness, Self-Respect, and Freedom,” in Ethics

Any reproduction is immoral: See David Benatar, Better Never to Have Been: The Harm of Coming into Existence (Oxford University Press, 2006).


Chapter 19: How should we build a virtual society?


The gamer’s dilemma: Morgan Luck, “The Gamer’s Dilemma: An Analysis of the
Arguments for the Moral Distinction between Virtual Murder and Virtual Paedophilia,”

**Virtual theft:** Nathan Wildman and Neil McDonnell, “The Puzzle of Virtual Theft,”


**VR analog of Milgram’s experiment:** Mel Slater, Angus Antley, Adam Davison, David Swapp, Christoph Guger, Chris Barker, Nancy Pistrang, and Maria V. Sanchez-Vives, “A Virtual Reprise of the Stanley Milgram Obedience Experiments,” *PLOS ONE*, https://doi.org/10.1371/journal.pone.0000039.


“Nasty, brutish, and short”: Thomas Hobbes, Leviathan i. xiii. 9.


Artificial scarcity. An extreme form of artificial scarcity arises with nonfungible tokens (NFTs) attached to digital artworks and other digital objects through blockchain technology. Some people pay large amounts of money for an NFT even it brings no obvious utility over and above being identified as the owner of the NFT. Here it appears that scarcity is being valued for its own sake. This form of artificial scarcity with no functional utility almost by definition applies only to luxury goods. However, less extreme forms of artificial scarcity for useful goods are only to be expected in a market system.


Chapter 20: What do our words mean in virtual worlds?

*Coffeehouse Conversation:* Douglas R. Hofstadter, “A Coffeehouse Conversation on the Turing Test,” *Scientific American*, May 1981. Reprinted in *The Mind’s I: Fantasies and Reflections on Self and Soul*, eds. Daniel C. Dennett and Douglas R. Hofstadter (Basic Books, 1981). Hofstadter develops this simulation realism further in a discussion of “SimTown” and “SimBowl” in *Le Ton beau de Marot* (Basic Books, 1997), 312–17. He also expresses a sort of virtual realism in discussing a virtual world of blocks on a table used by the AI program SHRDLU (p. 510): “However, whether the table was substantial or ethereal was of little import, since what really mattered was the patterns of objects in the situations, and those patterns were not in the least affected by their tangible physical existence or lack thereof.”

Gottlob Frege: See Michael Beaney’s *The Frege Reader* (Blackwell, 1997).


Limits to externalism: Tyler Burge, “Individualism and the Mental,” *Midwest Studies in Philosophy*, 4:1, pp. 73-122 (1979), has argued that the meaning of any term—even “seven”—can be “outside the head” of many speakers, when those speakers defer to others in their community. I’m setting aside this sort of social externalism by assuming the speakers are experts and don’t defer to others with regard to meaning.


If Sim Putnam says “I’m in a computer simulation”: In “Skepticism Revisited: Chalmers on The Matrix and Brains-in-Vats,” *Cognitive Systems Research* 41 (2017): 93–98, Richard Hanley suggests that if beliefs like “I’m not in a simulation” are false in a simulation, simulations may be skeptical scenarios after all. My response is that (as acknowledged in chapter 6) we may have some false theoretical beliefs about matters like this in a simulation, but that this does not lead to skepticism about ordinary beliefs.


Chapter 21: Do dust clouds run computer programs?


*First programmable electronic computer*: Other programmable computers developed around the same time as the Colossus include Konrad Zuse’s electromechanical Z3 machine, which was completed in Berlin in 1941 but never put into operation, and John
Mauchly and J. Presper Eckert’s ENIAC, completed in Philadelphia in 1945, which appears to have been the first working universal computer in Turing’s sense (Colossus and Z3 were programmable but not universal). Colossus remained secret for decades, and Z3 was destroyed during the war, so ENIAC was by far the most influential for subsequent computers. The first stored-program computers (including a modified version of ENIAC), which are closest to Turing Machines in that programs are treated as data, were developed in 1948. Turing’s mathematical work had little direct influence on the original design of ENIAC, but it may have played a role subsequently via the mathematician John von Neumann, who was influential in designing stored-program computers. See e.g., George Dyson, *Turing’s Cathedral* (New York: Pantheon, 2012); B. Jack Copeland, ed., *Colossus: The Secrets of Bletchley Park’s Codebreaking Computers* (Oxford UK: Oxford University Press, 2006); and Edgar G. Daylight, “Towards a Historical Notion of ‘Turing—the Father of Computer Science,’” *History and Philosophy of Logic*, 36:3, pp. 1-24 (2015).

---

*000*  

*000*  

*000*  
**In the absence of time:** A few philosophers (e.g., Sam Baron & Kristie Miller, “Causation in a timeless world,” *Synthese*, 191:12, 2867-86, 2014) have argued that you can get causation and counterfactual structure without time. If they’re right, you just might be
able to get computational structure without time, too, and maybe implement the Game of Life in a timeless world. But one still needs a timelike dimension for this to work—call it quasi-time—and implementing Life processes will require some sort of continuity of objects across quasi-time. All that is missing in the Dust.

*000 For a million generations or so: In a countably infinite dust cloud (one dust particle for every positive integer), we will expect to find particles that are always on for any finite period, but we can’t expect to find particles that are always on forever. If the dust cloud has the size of the continuum (one dust particle for every real number between 0 and 1), then we can expect to find particles that are on forever.

*000 The right pattern of counterfactuals: In correspondence, Greg Egan has expressed skepticism that satisfying counterfactuals could be crucial to consciousness: What matters is just the state of a system and not the counterfactuals. For example, if a neuron in the human brain is replaced by a recording that fires in exactly the same way in a given history (but wouldn’t do the same thing in counterfactual histories), this should yield the same state of consciousness.

I’m not sure about a single neuron: There’s enough redundancy in the brain that perhaps a few neurons could be replaced by recordings without degrading consciousness. But I’m confident that if my whole brain were replaced by a recording—a detailed movie of what all the neurons do in a given history, say, or a set of neurons programmed to fire in sequence—this would not yield consciousness like mine. The recording system wouldn’t be doing any information processing and isn’t sensitive to anything. If we gradually replace neurons by recordings, the capacities of the system will gradually diminish, and so will the associated state of consciousness. Admittedly, I don’t know
exactly how and why consciousness depends on counterfactuals and causal structure, but this is just part of the problem of consciousness: We don’t know how and why consciousness depends on anything.


Minority view among philosophers: In the 2020 PhilPapers survey, 54 percent of philosophers accepted or leaned toward a non-Humean view of laws of nature, which holds that laws (such as the law of gravity) involve more than regularity; 31 percent accepted or leaned toward a Humean view where laws are a matter of regularities. It’s plausible that the distribution of views about causation would be similar.

Chapter 22: Is reality a mathematical structure?


It sometimes happens that one physical theory makes another true: I’m not offering a
general analysis of when one physical theory makes another true, which depends on many subtle issues about the precise structural contents of these theories. One puzzle case arises from the so-called “holographic principle” and the associated AdS/CFT correspondence (see Leonard Susskind and James Lindesay, *An Introduction To Black Holes, Information And The String Theory Revolution: The Holographic Universe* [World Scientific, 2005]) in which certain higher-dimensional string theories (e.g., on the three-dimensional interior of a sphere) appear to be mathematically isomorphic to certain lower-dimensional quantum theories (e.g., on the sphere’s two-dimensional surface). I discuss the holographic principle and its connection to the simulation hypothesis in an online note.

*Argument that starts with structuralism and ends with simulation realism:* I’ve discussed predecessors to my simulation realism in work by Bouwsma, Hofstadter, Putnam, Davidson, and others. Where the use of structuralism to respond to external-world skepticism is concerned, I’ve found one predecessor: a paragraph by the philosopher of physics Lawrence Sklar in his 1982 article “Saving the Noumena” (*Philosophical Topics* 13, no. 1). Sklar entertains the idea that “the brain-in-a-vat account of the world is really equivalent to the ordinary material object world account, so long as the brain-in-a-vat account is suitably formally structured” (p. 98), but immediately dismisses the idea as being too close to instrumentalism.

*It does not address the problem of other minds:* Grace Helton, in “Epistemological Solipsism as a Route to External-World Skepticism” (*Philosophical Perspectives*, forthcoming), and in other work on structuralism and skepticism, argues that if others don’t have minds, then many ordinary physical objects don’t exist, including social
entities such as cities, churches, and clubs that depend on minds for their existence. If so, a structuralist anti-skeptical strategy that does not establish the existence of other minds does not establish the existence of social entities, and skepticism about the social realm remains open. Still, I think it is plausible that atoms, cells, trees, planets, and other physical objects do not depend on other minds for their existence. If so, skepticism about other minds does not lead to skepticism about the ordinary physical world.

*Reminiscent of Kant’s transcendental idealism:* In my “The Matrix as Metaphysics” (2003): “One might say that if we are in a matrix, the Kantian ding-an-sich (thing in itself) is part of a computer-an-sich!” Barry Dainton’s “Innocence Lost: Simulation Scenarios: Prospects and Consequences” (2002, https://philarchive.org/archive/DAIILSv1) also suggests connections between the simulation hypothesis and transcendental idealism (“In Kantian terms, virtual worlds of the communal variety are *empirically real*, even if *transcendentally ideal*”), as does Eric Schwitzgebel in “Kant Meets Cyberpunk,” *Disputatio* 11, no. 55 (2019): 411–35.


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


The two images: It’s tempting to identify the manifest image with Kant’s realm of appearances and the scientific image with Kant’s realm of things in themselves. But this would not be quite right. Both the manifest image and the scientific image are knowable, so Kant would count both as part of the realm of appearance. If we combine the frameworks, we have three stages: the manifest image, the scientific image, and things in themselves. The scientific image is perhaps a step from ordinary appearances toward things in themselves. Indeed, at certain points Kant talks about a distinction between appearances and things in themselves in the “empirical” realm (as opposed to the transcendental realm). Kant’s empirical things in themselves seem not far from the objects of Sellars’ scientific image.

Colors exist only in the mind: In *Il Saggiatore* (*The Assayer*, 1623), Galileo writes, “[T]hese tastes, odors, colors, etc., so far as their objective existence is concerned, are nothing but mere names for something which resides exclusively in our sensitive body (*corpo sensitivo*), so that if the perceiving creatures were removed, all of these qualities would be annihilated and abolished from existence,” in *Introduction to Contemporary Civilization in the West*, 2nd edition, vol. 1, trans. A. C. Danto (Columbia University Press, 1954), 719–24.

Spatial functionalism: I introduced spatial functionalism in chapter 7 of *Constructing the World* (Oxford University Press, 2012) and developed it further in “Three Puzzles about


*Hoffman’s case against reality:* Online appendix.


Chapter 24: Are we Boltzmann brains in a dream world?

What if God created reality five minutes ago: Bertrand Russell, The Analysis of Mind (George Allen & Unwin, 1921), 159–60.

Temporary simulation skepticism: In defending Hilary Putnam’s externalist response to skepticism, Joshua Rowan Thorpe (“Closure Scepticism and the Vat Argument,” Mind, 127:507, pp. 667-90, 2018) argues against temporary skeptical scenarios on the grounds that we can use our secure knowledge of the past to rule them out. For example, if the past was the way we think it is, then nobody had the perfect-simulation technology to put us into a perfect simulation. So we at least need to independently question our knowledge of the past to justify this sort of skepticism. One response for the skeptic is to postulate a dual-simulation scenario, in which the simulators moved us from one simulation to another five minutes ago. In that case, our ordinary knowledge of the past may be mostly justified, but we certainly didn’t know that nobody had perfect-simulation technology.

A related thought: Once advanced simulation technology becomes rife, then the hypothesis that we’re plugged in to a perfect simulation may become a serious one, entirely consistent with our knowledge of the past. At that point, it may be hard to know we’re not in a temporary simulation. But that’s not the situation we’re currently in.


God is playing the role of the computer: Peter B. Lloyd (“A Review of David Chalmers’ Essay, ‘The Matrix as Metaphysics,’” 2003, DOI:10.13140/RG.2.2.11797.99049), who responds to my analysis from a Berkeleyian idealist perspective, suggests that even Berkeley’s God might be running some sort of shortcut (just-in-time) simulation in order
to be more economical.

*000*  *When Zhuangzi dreams of the butterfly, there is a real dream butterfly:* Zhuangzi’s own discussion has an element of virtual realism, insofar as it stresses the reality of both the butterfly and Zhuangzi (though unlike my analysis, Zhuangzi’s analysis also stresses the distinction between the butterfly and Zhuangzi). See Hans Georg Moeller, *Daoism Explained: From the Dream of the Butterfly to the Fishnet Allegory* (Open Court, 2004).

*000*  *Novels and other fictions:* Online appendix. What about novels and other fictions? Do events in these really take place in the head of the author or reader? I would say usually not. A reader’s mind will not usually have anything like an interactive world-model. An author’s mind may contain more of a model, but in many cases the model may often be more like a script building toward an outcome than a genuine open-ended and interactive world. For some authors in some cases, writing a novel may unfold as a full-scale interactive simulation. In that case, the events could have at least the limited mind-dependent reality of the events of a dream.

Interactive novels are a special case. In most existing interactive novels, the interaction is too intermittent for this to involve anything like a virtual world. However, a highly interactive novel would approach something like a text adventure game. *Colossal Cave Adventure* involves a genuine virtual world: It’s interactive and computer-generated, with its state encoded in a database of virtual objects, even though it’s not immersive. Someone playing *Colossal Cave Adventure* is genuinely interacting with a virtual world. The same goes for the virtual worlds involved in games such as *Dungeons and Dragons*, which are traditionally realized in the notebooks, props, and memories of participants; see Jon Cogburn & Mark Silcox, eds., *Dungeons and Dragons and Philosophy* (Chicago:
Open Court, 2012). Even if there’s no computer here, there’s something akin to a virtual world.

Ordinary interactive novels and games don’t really raise a skeptical issue. We can plausibly know that we’re not in an ordinary Dungeons and Dragons game, since those games would not support our detailed perception. One could perhaps make a case that we’re in an unusually rich version of the game that models much of our perception and the physical world. But this brings us back to a more standard version of the simulation hypothesis.

*000 Experiences not generated by the external world: Here’s one more empty-world hypothesis. Let Ordinary World be a world like ours. Then let Experience World be a world containing only states of consciousness, with one law of nature: The states of consciousness in Experience World at time $t$ are the same as those in Ordinary World at time $t$ (where Ordinary World is specified by its laws of nature and initial conditions). Then beings in Experience World will have experiences just like Ordinary World, but there will be no external world there.

To respond: I’m not sure that there could be a law of nature like this. If there can be such laws, they’re certainly more complex than the laws of Ordinary World, so there’s a simplicity case against the hypothesis that we’re in Experience World. I’d also argue that for this law to work, Experience World needs states that reflect the states of Ordinary World. Once we have those, Experience World is no longer a world with just conscious states; it’s a world where conscious states interact with an external world.

In Mueller’s ingenious framework, observations evolve from other observations by a single law in algorithmic information theory. Roughly: The probability of the next observation being A is the algorithmic probability of A given earlier observations, which is determined by the length of the shortest algorithm that produces previous observations and A. It’s highly unlikely that Mueller’s framework would produce even the appearance of an external world, as opposed to a regular parade of internal experiences. In any case, perhaps one could argue along the same lines as for Experience World: Using algorithmic probabilities in a law of nature requires use of the relevant algorithms that will then support an external world.


Only a tiny minority would have ordered experiences: This is a form of the typicality reasoning I raised questions about (in discussing Carroll’s arguments against the simulation hypothesis) in the notes to chapter 5. Still, it’s reasonable to give typicality reasoning more weight in this case, as the relevant numbers are so enormous. If we’re in a Boltzmann-friendly universe, only the tiniest minority of conscious beings will have ordered experience. In Carroll’s earlier arguments, there’s a case that a small minority of conscious beings won’t be in bottom-level simulations, but the scale is very different (one in a thousand, say, compared to one in $10^{120}$ or higher).
Index