2 Knowledge and skepticism

Red pill or blue? In the eponymous groundbreaking film, Neo takes the red pill and exits the "matrix." Once out of the matrix Neo realizes he had been living his life in a virtual reality world created by **artificial intelligences**. Of course, people who've not yet taken red pills don't know that they inhabit a virtual reality world. It appears real to them . . . just as our present world appears real to us.

The life to be lived within the matrix isn't bad at all. Residents live in nice houses and have interesting jobs. They eat delicious meals—steak is a favorite. They read books, watch television, and listen to music. They take trips to New York, have sex and play volleyball. Residents in the matrix have interests and beliefs identical to ours.

When considering alternatives open to people in the matrix, life looks comfy. If you take the red pill, you wake up rudely and you thrash around in an amniotic sac. After choking on the fluid, you realize you have no idea where you are. You occupy an isolated pod, one amongst millions, set deep underground. Someone might pick you up to enlist you in the fight against the AI's, but this means you'll be serving on a cramped battle craft, always on alert for preying AI attack vessels. You'll lose all your matrix-family. Your meals will be nutrimeat and yeast cultures. You'll never sit in a Lazy Boy, eat a pizza and watch Direct TV again.

So would you really—*really*—want to take the red pill?

The ancient philosopher Plato offers a clear answer to this question in his "Allegory of the Cave" (see §2.11). If you knowingly don't take the red pill, you are living a lie. You believe that your matrix-steak came from a cow, but it didn't; that you are going to sleep in your room, but you don't have one; that your spouse loves you, but you don't have a spouse. Instead, you wallow in that fluid-filled sac all alone.

Whether you want to be in the matrix or not, you probably can't know whether you are in one now or not. Of course, we *think* we know we're in the real world. However, a number of philosophers, upon investigating the nature of knowledge, have concluded that we know very little, if anything. What can we know?

After completing this chapter, you should be able to do the following:

- Explain **theism** and **atheism**, the concept of God and the **paradox of omnipotence**
- Explain and assess the **design argument** for God's existence
- Explain typical problems with **analogical arguments**
- Explain the concept of evil and its variations, and explain and assess the **argument from evil** against God's existence
- Explain and assess several criticisms of the argument from evil
- Understand the problems surrounding the relationship between proofs for God's existence and knowledge of God's existence.

2.0 THE CONCEPT OF KNOWLEDGE

The question "What can we know?" can't be answered fully until a prior question is answered, namely, "What is knowledge?" Though the concept of knowledge seems clear, *epistemologists*—philosophers of knowledge—find that giving an adequate definition is no easy task. "Knowledge" has many uses and many sources. Jones knows his next-door neighbor, Smith; knows *that* Smith is a reservist in the National Guard; knows *what* Smith does on Saturdays; knows *how* to annoy Smith; and knows *why* Smith is a social worker rather than a programmer or a server.

The primary concept of knowledge is knowledge *that* some statement is true. Many uses of the concept can be put in terms of knowledge *that*. Jones' knowing *what* Smith does on Saturdays is nothing other than Jones' knowing *that* she reads home improvement books on Saturdays. "Knowledge that" is **propositional knowledge** since what you know is a statement that expresses a proposition about something. So epistemologists address the question "What is knowledge?" by answering the question "What is propositional knowledge?"

What conditions must someone fulfill to have propositional knowledge? Something like the following three conditions must be met. For an **agent** A, Agatha, to know some statement P,

(2.1) P must be *true*

(2.2) A must *believe* P, and

(2.3) A must be *justified* in believing P.

We have already defined *truth* as correspondence to reality. *Belief* is simply one's mental assent to a statement. Knowledge requires that condition 2.1 is appropriately related to condition 2.2. Not just any true belief deserves the honorific title of knowledge,

after all. *Justification* refers to evidence for a belief that increases the likelihood that the belief is true.

This characterization of justification distinguishes A's *justification* for believing P from A's *rationality* in believing P. There is an extremely important difference between the two. If A is justified in believing P, then A's belief P is rational. But the converse is not true. If A is rational in believing P, it does not follow that A is justified in believing P. "Rationality" refers to the conditional probability of a belief on the basis of background beliefs. Suppose I believe that aliens will attack Earth tomorrow because I also believe that I met the aliens last year, and I believe that they told me many true things, including things about this attack. Since these are my background beliefs, I am rational in believing that aliens will attack tomorrow.

However, I can be rational in this belief even if none of these background beliefs are true and even if these background beliefs are themselves unjustified by further beliefs. The mere fact *that I believe* that I met aliens last year and that they told me these things doesn't evidentially support—doesn't make it likely to be true—that aliens will attack tomorrow. To be *justified* in believing that the aliens will attack tomorrow, I need background beliefs that offer evidential support for my belief in the attack by actually increasing the likelihood that my belief in the attack is true. (See §2.0.)

Philosophical definitions of the key term "justification" oscillate between two different meanings. On one theory, **internalism**, justification takes the form of evidence or reasons possessed by the agent. This type of justification converts true belief into knowledge. Agatha knows P because *she herself* has *reasons to believe* P. Internalism states that justification is constituted by certain kinds of mental states internal to A, which provide A with a reason to think P is true. Naturally the internalist must say much more about justification than we have on his behalf, but this is sufficient to contrast the view with its major alternative, **externalism**.

We can characterize externalism as a theory of justification on which justification depends upon the features that produce the belief, rather than on reasons possessed by the agent. On one externalist proposal the reliability of the process that causes belief P is what justifies belief P for person A. Thus A's belief P is justified, on the "reliabilist" form of externalism, only if P is produced by a reliable process. A *reliable* belief forming process is a process that would produce more true beliefs than false beliefs over the long run. According to reliabilist externalism, A's perceptual beliefs will be justified because perception is a reliable belief-forming process. A's guesses or A's beliefs based upon her horoscope will not be justified on this theory because guessing and astrology do not produce more true beliefs than false beliefs. One benefit of externalist theories is that we can be justified to belief many widely held beliefs. The standards of internalist theories are higher, but they yield smaller quantities of known beliefs.

One alleged problem with externalism is that it doesn't take philosophical problems of knowledge as seriously as does internalism. One often-quoted understanding of internalism comes from W. K. Clifford. He says, "To sum up; it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence."[1] Internalism recommends—in Clifford's version, demands—that individual believers undertake a quest

to justify their beliefs with evidence that is accessible to the believers themselves. But externalism threatens the internalist's picture of the philosopher deep in critical thought about the problems of knowledge. Some venerable problems in epistemology are solved too easily on externalism. If externalism is true, then whether Agatha thinks critically about the sources of the reliability of her belief bears little relation to her justification for it. When confronted with a skeptical scenario—she might be in a simulation of reality—the internalist's fundamental question is: "What reasons has she to justify her belief that she is not in a virtual reality environment?" The externalist's fundamental question is: "Are her beliefs generated by cognitive processes that produce more true beliefs than false beliefs?"

Science fiction stories often exemplify externalism by enhancing the minds of their characters, whether through technology, pharmacology or through some super-special "psionic" power. The award-winning short story called "Flowers for Algernon" explores externalism through a drug treatment given to the protagonist, Charlie Gordon. Charlie, a mentally retarded janitor, volunteers for an intelligence-enhancing drug, which causes a steep rise in his intelligence. At points in the story, Charlie can't understand how he knows that he knows, even though his beliefs are true. Daniel Keyes, the author, writes the story in the first-person, deftly conveying the rise and the precipitous fall in Charlie's intelligence as the drug's effects fade. (*Rental*: *Charly* (1968) stars Cliff Robertson, who won the Academy Award for Best Actor.)

In the Philip K. Dick short story "Minority Report," set in 2054, three individuals possess the ability to foresee glimpses of the future. (*Rental*: *Minority Report* (2002), directed by Steven Spielberg and starring Tom Cruise.) These three "pre-cogs," as they are called, glimpse images of the near-term future in which citizens are in imminent danger of losing their lives—this seems to be the only information that their pre-cognition faculty gives them. The sources of their incredibly valuable foreknowing ability remain somewhat mysterious. But this belief-forming faculty works. Mostly. The criminal justice system of the future uses this special form of knowing to its advantage; Dick imagines a division of the police force devoted to "pre-crime." These the images are translated into data and fed to the police information technology system, which automatically dispatches teams to intercept the individual about to commit murder.

By using this exotic belief forming faculty, do these three people *know* those belief? The reliabilist externalist would answer "Yes" because it is sufficient for knowledge that the belief is true and is caused by a reliably truth-conducive faculty of the mind. Someone might reply with any number of objections, including: Scientists do not know how their exotic belief-forming faculty works so well. The pre-cogs themselves don't understand how they have come to have this power. The origins of this faculty appear in tension with standard evolutionary theories about the origins of the components of our brains. Only three people in the world appear to have this special capacity. Taken together, these objections go some way toward undermining confidence in our judgment about the faculty.

These objections are voiced from an internalist perspective, one according to which we need more information about belief-forming faculties before we give them our credence. For example, we need a larger test group to determine that the very small results are not a statistical fluke. But the externalists will insist that, despite the many things that

we do not know about the faculty, the most important fact we do know is that it works reliably to produce more true than false beliefs.

As described, this scenario raises a host of epistemological questions, and the life-and-death setting of this gadget story forces the characters to make difficult ethical decisions based directly upon the justificatory status of their beliefs about what will happen. But Dick puts a further twist in our epistemic intuitions about the faculty of pre-cognition when he reveals the meaning of the title. In some cases the triumvirate of pre-cogs fail to reach unanimity about what will happen. When there is disagreement, the "minority report"—the one report of three that disputes the beliefs formed by the other two—is given careful analysis. The story unfolds into an internal affairs mystery about corrupt officers who have tainted the minority reports in some key investigations.

BOX 2.A: CLAIRVOYANCE AND EXTERNALISM

Externalism attempts to define "knowledge" in a way that implies that the justification for a known belief can come in the form of its connections to the facts that make it true. On reliabilist externalism, if the process that produces the belief is reliable, then it is known. In contrast, the internalist says that the justification for a belief arises from the reasons that a person has to believe that it is true. This difference gives rise to a line of criticism against externalism that can be illustrated with the help of science fiction examples.

In the re-imagined *Battlestar Galactica* series, President Laura Roslin battles her terminal illness through a psychotropic remedy, "chamalla." After taking chamalla, Roslin experiences hallucinations that she interprets as prophecies about the future, which are given to her by the gods. The beliefs induced by chamalla are more often true than false. This meets the reliabilist externalist's criterion for justified belief. (*Rental*: *Battlestar Galactica* season one episodes "Flesh and Bone," "The Hand of God," and "Kobol's Last Gleaming, Part I" all feature Roslin making important decisions based in part or whole upon chamalla-induced visions.)

So, is Roslin justified in believing them? Within an epistemological context, Laurence BonJour's thought experiments convinced many philosophers that externalism was false. This is one of his cases:

> Samantha believes herself to have the power of clairvoyance, though she has no reasons for or against this belief. One day she comes to believe, for no apparent reason, that the President is in New York City. She maintains this belief, appealing to her alleged clairvoyant power, even though she is at the same time aware of a massive amount of apparently cogent evidence, consisting of news reports, press releases, allegedly live television pictures, etc., indicating that the President is at that time in Washington, D.C. Now the President is in

> fact in New York City, the evidence to the contrary being part of a massive official hoax mounted in the face of an assassination threat. Moreover, Samantha does in fact have completely reliable clairvoyant power, under the conditions that were then satisfied, and her belief about the President did result from the operation of that power.
>
> (BonJour, 1980: pp. 59–60)

According to the externalist, Samantha knows that the President is in New York City. Not only does she not have a reason to believe this, but she also has reasons against this. Does she know?

We will be working with an internalist account of justification in the remainder of the chapter, though we will say more about this controversy in our conclusion. Meantime, though, one needn't take a stand on this issue to see amiss in our definition of knowledge.

This three-part, "**justified true belief**" definition of knowledge held sway for ages. But Bertrand Russell and, later, Edmund Gettier, devised clever counterexamples to it—"Gettier cases." Imagine that every day for a decade Agatha has passed by a grandfather clock on the landing of her stairway. She has many true and well justified background beliefs, including: the clock has told time accurately for the past decade; my eyes are in great working condition, allowing me to read the clock face; I know how to tell time; etc. These background beliefs are all true, she has good evidence for thinking that these beliefs are true and the set of her background beliefs makes it likely that her belief that it is 10 a.m. is also true. One morning Agatha looks at the grandfather clock, sees that its hands indicate that it is 10 o'clock. Unbeknownst to her, the clock stopped precisely twelve hours earlier. But it so happens that it is 10 a.m. at the moment she forms the belief that it is 10 a.m. Her evidence for her belief is impeccable and it does make it likely to be true that it is 10 a.m. so her belief that it is 10 a.m. is justified.

However, we are reluctant to attribute to Agatha *knowledge* that it is 10 a.m. Something seems amiss. It is only by a serendipitous accident that Agatha happened to glance at the clock just when she did, which happened to be exactly 12 hours from the time the clock happened to stop working. Her belief depends on a number of coincidences; she got very lucky this time. This Gettier case and others like it have convinced many people of the deficiency of the justified, true belief definition of knowledge.

The creation of Gettier cases prompted the search for a "fourth condition" for knowledge. Added to the previous three, a fourth condition involves specifying just what is missing from the JTB account. Epistemologists disagree widely about the content of the fourth condition. One such condition might read:

(2.4) There is no true statement that would render A's belief P unjustified, were A to be aware of that true statement.

In this Gettier case, Agatha's justification for the belief that it is 10 a.m. depends on her belief that the clock works well. Though it has in the past, it doesn't now work well. We can call 2.4 a *no false belief condition* for it specifies that Agatha will have knowledge provided she meets the first three conditions, *and* her justification for the belief doesn't depend on a false belief.

In the thought experiment under discussion Agatha's circumstances fail to meet condition 2.4. The true proposition that defeats Agatha's justification for believing that it is 10 a.m. is "The clock is not functioning." Were Agatha to be apprised of this fact, she would no longer be justified in believing that it is 10 a.m. Since she is not justified in believing proposition P, she does not have knowledge of P.

BOX 2.B: GETTIER CASES AND THE DEFINITION OF KNOWLEDGE

In order to gain a better understanding of the JTB definition of knowledge, it is helpful to create cases of purported knowledge in which the agent believes statement P, P is true and A is justified in believing P, but in which A does *not* know P. Here is the most famous such case, created by Edmund Gettier.[2]

Suppose that Smith is justified in believing the belief "Jones owns a Ford," even though this belief is false. (Imagine that Smith saw someone in Jones' driveway who looked exactly like Jones washing a Ford.) Smith draws the valid inference that "either Jones owns a Ford, or Brown is in Barcelona." This is a valid inference because if you know P is true, then you also know that either P or Q is true.

Smith does not have any evidence to think that Brown is actually in Barcelona, but by sheer luck Brown is in Barcelona. So Smith's belief is true because it is true that "either Jones owns a Ford, or Brown is in Barcelona." Furthermore, Smith has evidence for this disjunctive statement—his original evidence that Jones owns a Ford. So on the JTB definition, even though he has inferred it from a false belief, Smith *knows* that "either Jones owns a Ford, or Brown is in Barcelona." But surely Smith doesn't really know this—it is by a stroke of luck that he is justified in believing it.

2.1 SKEPTICISM

What we have said about the term "knowledge" in 2.0 assists us in understanding epistemological challenges to knowledge claims. For every kind of knowledge claim that has been offered, it seems there has been a skeptic who attempts to argue that the claim fails to meet the requirements for knowledge. A skeptic who denies that we can have any knowledge is called a *global skeptic*.

The knowledge claims that skeptical arguments target include claims

- about religious and other supernatural matters
- about the thoughts, feelings, and sensations of other people
- about ethical and moral judgments
- about memories
- about the laws and makeup of the world around us
- about abstract principles, such as the laws of logic and mathematics
- about ourselves and our own thoughts, feelings, and characteristics.

The global skeptic holds that we have no knowledge of statements within any of those topics. Someone who holds that we lack knowledge of statements in some but not in all of the topics above is a *local skeptic*. For example, the story included in this chapter, Philip K. Dick's "We Can Remember It for You Wholesale," portrays one man's struggle with a form of local memory skepticism. The protagonist tries to identify which of his memory beliefs are truthful and which are not, but in the process he becomes skeptical of just about all his memories.

In denying that we have knowledge, the skeptic does *not* deny that some beliefs are better than others. Even the most extreme global skeptic can agree that we are *less justified* to believe that the moon is made of green cheese than to believe that it is made of rock.

BOX 2.C: ORSON WELLES, MARTIANS AND UNJUSTIFIED BELIEFS

Orson Welles was responsible for what must be one of the biggest practical jokes of the twentieth century. On October 30, 1938, under his direction, CBS radio interrupted the broadcast of an orchestral concert to alert listeners to the landing of a Martian spacecraft in New Jersey. (The description of the invasion was based largely upon the book *War of the Worlds*, authored by H. G. Wells.) The number of people convinced that a Martian invasion was taking place is often blown out of proportion to the point that people believe that the majority of listeners were duped. (There were, after all, several announcements that the broadcast was a radio play, not a newscast.) Nonetheless many acted as though they believed it by driving hundreds of miles to be with loved ones and holding impromptu religious services. Were they epistemically justified in believing there was an invasion?

When this stunt was repeated in February 1949, in Quito, Ecuador, a larger portion of the listening audience was fooled. This time listeners became so angry at having been deceived that a violent horde of people began rioting in the city and burned to the ground the radio station that broadcast the play.

Rental: Tom Cruise stars in a remake of H. G. Wells' *War of the Worlds* (2005).

The skeptic attacks the third condition of the above analysis of knowledge, statement 2.3. The skeptic allows that your belief may be rational for you, in the sense of rationality defined in §1.3. But the skeptic denies that you have sufficient evidence to be justified, so the skeptic will conclude that you do not have knowledge.

In the seventeenth century René Descartes added a number of arguments for and against several forms of skepticism in his important work *Meditations on First Philosophy*. His goal was not to wreak havoc in the then-current belief systems, even though the *Meditations* had that effect. Instead, Descartes wanted to give our knowledge claims a trial by fire, putting them through rigorous tests so that they would be better able to withstand skeptical attacks.

In the first of Descartes' *Meditations* (see reading §2.12) he creates a succession of arguments for skeptical positions. His skeptical considerations successively increase in scope, so that each skeptical argument calls into doubt more beliefs than the prior argument. Descartes starts by noticing that the propositions he had devoutly believed in the past have turned out to be false. He decides to discard any belief that is based on any evidence that is less than certain. If the evidence is less than certain, the conclusions drawn from the evidence cannot constitute knowledge, for they too will be uncertain.

Descartes realizes that much of what we believe is based on sense experience. We make judgments based on what we see, hear, feel, and so forth. Each of us, however, has experienced some form of sensory illusion.

BOX 2.D: BLIND SPOTS

So you don't think you need to worry about perceptual skepticism because your sensory systems work well? Ever considered the possibility that your sensory systems are systematically causing false perceptions? Let us help: put your right hand over your right eye. With your face about two feet away from the page, stare with your left eye at the heart and slowly move toward the book.

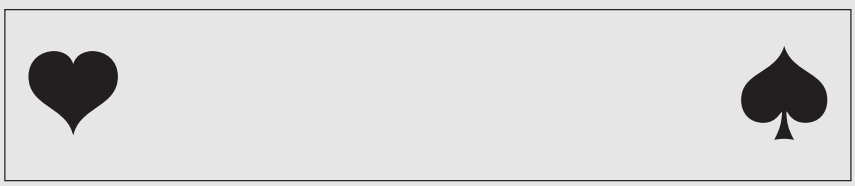

Figure 1 Heart and spade

The spade should disappear at some point and be replaced by a white background of the same color that the page is printed on. You have found your blind spot.

At the back of each eye is the location at which an optic nerve enters the retina. This blind spot has no light receptors, therefore no light is registered by your eye and brain as being present in the location in your field of vision directly opposite your blind spots. Therefore, no light is picked up there. The reason you do not detect it normally is because both eyes working together help to minimize the effects of the blind spot. Your brain deceives you by filling in the blind spot with the colors nearest the blind spot.

When we experience an illusion, we form a false belief about the way the world is because our perceptions do not correspond to the world. I see what appears to be a body of water across the desert highway, but on approach I realize that there was no body of water and that, instead, my eyes were playing tricks on me. Descartes offers another example: from across the fields, I see what appears to be a round tower but closer in I see that it is square. Descartes infers that, since his senses sometimes deceive him, he is not justified in believing anything his senses tell him. In his words, "Whatever I have up till now accepted as most true I have acquired either from the senses or through the senses. But from time to time I have found that the senses deceive, and it is prudent never to trust completely those who have deceived us even once" (Meditation I, Descartes, 1984: p. 12).

Descartes provisionally endorses this skeptical argument about the senses. But in the cultural and historical context of his time, Descartes wants to clear away all the opinions he accepted blindly upon testimony. In the early 1600s, when Descartes was thinking about these problems, the scientist Galileo was condemned by the Catholic Church twice because Galileo dared assert that the Sun was in the center of the solar system, not Earth. The Catholic Church exerted considerable control over peoples' beliefs, and it had taught false beliefs about science and the world. Descartes wanted a fresh start for epistemology.

This provides Descartes his motivation for using what he calls the "Method of Doubt." He illustrates the method with a simile. Our set of beliefs is like a barrel of apples, some good, some rotten.

> Suppose [someone] had a basket full of apples and, being worried that some of the apples were rotten, wanted to take out the rotten ones to prevent the rot spreading. How would he proceed? Would he not begin by tipping the whole lot out of the basket? And would not the next step be to cast his eye over each apple in turn, and pick up and put back in the basket only those he saw to be sound, leaving the others?
>
> (Replies to the Objections to the Meditations, Descartes, 1984: p. 324)

One rotten apple can spoil the whole barrel of them, so the best way to prevent that from happening is by tipping over the whole barrel. In much the same way, one false perceptual

belief can be used as a basis for many other false perceptual beliefs. Descartes recommends that we reject all the beliefs produced by some belief-forming process—in this case, vision—as being methodologically false.

2.2 DREAM SKEPTICISM

By using skeptical arguments, Descartes intends to tip the barrel over and call into question as many beliefs as possible. But perceptual skepticism didn't go far enough and didn't cause him to doubt enough beliefs. For example, even if I can't know any perceptual beliefs, I can still be justified in beliefs based on testimony from others, beliefs based on my memories and beliefs based on mathematical reasoning. So, next, he broadens the scope of skepticism with another argument having to do with dreams.

In (at least some of) our dreams, our minds create vivid experiences that can seem real to us. In some cases they appear so real that they can be mistaken for waking experience. Because dream experience can appear indistinguishable from waking experience, the skeptic argues that any waking experience I seem to have might be dream experience instead. Descartes builds a skeptical argument from these considerations about dreams.

BOX 2.E: *STAR TREK: THE NEXT GENERATION*, "SHIP IN A BOTTLE"

In "Ship in a Bottle," (season 6, episode 12) the android Commander Data is involved in a role-playing, virtual reality detective game based on the Sherlock Holmes stories. Data commands the computer to create a criminal worthy of Data's detective abilities and it does so—with startling results. Captain Picard and Data are deceived by Moriarty, Data's computer-created, virtual reality opponent (and Holmes' literary nemesis). Moriarty deceives the two into believing that they are on the bridge of the *Enterprise* when in truth they are in a virtual reality holodeck simulation of a bridge created by Moriarty himself. Fear not: Picard and Data eventually realize the deception in time to save the ship and her crew. They escape by tricking Moriarty in the same way he tricked them.

For some time while in the simulation, Picard and Data believed that they were actually operating the real ship. Though they were in the holodeck, they did not know they were. Likewise, we have various beliefs when we are dreaming about the dream-scenarios. But while in a dream, we do not know that we are. The skeptic seeks to build an argument, inspired by these situations, for the conclusion that one cannot know that one is not now dreaming.

We'll put this argument in the first person for obvious reasons. Begin with a factual claim:

(2.8) There is a qualitative similarity between what seems to be my dream experience and what seems to be my waking experience.

This is a statement of the empirical fact that one's dream experience greatly resembles one's waking experience at least some of the time. Descartes draws some inferences from this observation.

(2.9) If (2.8) is true, then there are no certain indications by which I can distinguish what seem to be dream experiences from what seem to be waking experiences.

(2.10) If there are no such certain indications, then I cannot *know* I'm not dreaming right now.

(2.11) So, I do not know that I'm not dreaming right now.

So far this argument has the following valid structure:

(2.8) P

(2.9) If P, then Q

(2.10) If Q then not K

(2.11) So, not K

This conclusion, that I cannot know that I'm not dreaming right now, is the thesis of **dream skepticism**. But what does Descartes mean by "knowledge"?

According to Descartes, for agent A to know proposition P, A must believe P, P must be true and A must be justified in believing P. So far, so good. But Descartes has his own special definition of "justification." To be justified a belief must be *indubitable*; this term means "unable to be doubted."

This type of skepticism offers two possibilities: I am in the real world or I am in a virtual world? In the *Matrix* films, characters seem to know something exceedingly important: *they know that they are either in the matrix or in the real world*. This notion that there might be a *plurality* of worlds—many virtual worlds and one real one— instead of merely a duality of worlds is not explored well in the films. The characters are given a clear and simplistic choice between two and only two alternatives. This point crystallizes in the observation that none of the characters ever appear to take seriously the hypothesis that what they think of as their non-matrix world—in which they move around in their hovercraft and subterranean caves—*might itself be part of the matrix*.

Let's return to Descartes' conclusion in the dream argument: I cannot know that I am not now dreaming. At this point in Descartes' dream argument, we have a similar phenomenon going on. On common interpretations of Descartes' dream skepticism, he too

allows us to presume that we are either in a dream world or in the real world. There are only two options there, just as in *The Matrix*. Once they take the red pill, characters in the film are given knowledge of the difference between appearance and reality. For these people, they refute global skepticism by the simple act of swallowing!

Descartes takes the argument further. He realizes that if he can't know that he isn't now dreaming, then he also can't know that any of the perceptual beliefs he has are true. His perceptual beliefs—about what he sees, say—might instead be caused by computer programmers who are writing Descartes' simulation. They might be caused by Descartes' mental delusions.

> All these considerations are enough to establish that it is not reliable judgement but merely some blind impulse that has made me believe up till now that there exist things distinct from myself which transmit to me ideas or images of themselves through the sense organs or in some other way.
>
> (Meditation III, Descartes, 1984: p. 27)

In this passage Descartes is adding another obvious observation, and follows it with further conclusion:

(2.12) In dreams, the mind-independent physical objects I believe cause my experiences are not causing my experiences.

(2.13) If (2.11) and (2.12), then I cannot know that there is a mind-independent, physical world.

Since premises 2.11 and 2.12 are in place, the antecedent—the if–clause—of 2.13 is true. This implies 2.14:

(2.14) So, I cannot know that there is an external, mind-independent world.

This conclusion is stronger than the earlier claim that I can't know that I'm not dreaming now. This argument for dream skepticism has now called into doubt beliefs about the physical world due to premises 2.12 and 2.13. The scope of this conclusion is now much wider than 2.11. Now this argument is quite strong, but not strong enough.

2.3 EVIL DEMON SKEPTICISM

We mentioned that in the tumultuous cultural and historical context of the early seventeenth century, Descartes wanted to clear away as many previous beliefs and biases as he could in order to put philosophy and science on firmer footing. The dream argument created doubt about many more types of beliefs than did his earlier perceptual skepticism argument. However, dream skepticism still didn't create doubt about all beliefs. Descartes thought that even if the full dream argument is sound, some knowledge would be left

intact. Specifically, knowledge of all sorts of mathematical beliefs would remain even if he can reach the conclusion in 2.14. Agent A's belief that 1 + 1 = 2 will be known so long as A can entertain the meaning of the terms in the belief. And this is something that she can do in dreams (says Descartes). On this basis, Descartes sought—and found—an even more pervasive form of skepticism. He hypothesized that an all-powerful evil God might exist and deceive us all. In order not to offend religious authorities at the time, Descartes redescribes the skepticism as being caused by an evil demon or evil genius.

It seems possible that there is a terrifically powerful being that controls all our thoughts and experiences, and deceives us into thinking all sorts of things that are untrue. We might be deceived in our thoughts about basic abstract principles. Each time we add 1 + 1 we might be getting different results—and yet we might be manipulated in such a way as to be permanently unaware that we are getting different results. The evil demon could induce in us a strong feeling of regularity, consistency, and certainty in our abstract thought even when we were pitifully mistaken. The evil demon (or deceiving God) might be altering the basic metaphysics of the universe and structuring our thoughts to misrepresent it. Here is Descartes' own description of this thought experiment:

> [F]irmly rooted in my mind is the long-standing opinion that there is an omnipotent God who made me the kind of creature that I am. How do I know that he has not brought it ab out that there is no earth, no sky, no extended thing, no shape, no size, no place, while at the same time ensuring that all these things appear to me to exist just as they do now?
>
> (Descartes, 1984: p. 21)

BOX 2.F: INFLUENCES ON DESCARTES' EVIL DEMON ARGUMENT

Suppose you believe that demons can inhabit and control human beings. Demons do bad stuff, so it is only natural that people inhabited with demons would eventually be brought to trial. In the 1630s in a famous case in Loudon, France, a priest named Grandier was alleged to be possessed by a demon, and to have infested a convent of nuns with a legion of demons. This made headlines at the time, and Descartes was well aware of the trial and reflected upon the fascinating philosophical problems to which it gives rise.

Wouldn't Grandier's apprehension by the police be evidence against his being possessed, since demons are presumably so powerful that they could escape apprehension? Couldn't the demon control those around him so as to manipulate court proceedings of any kind? Could Grandier's own testimony ever be taken as veridical—even if he swore to tell the truth on a Bible?

Further reading: None other than author Aldous Huxley was so intrigued by this

trial that he wrote a book about it called *The Devils of Loudon* (New York, 1952). Huxley's interest in social deception and control arises in his much more famous dystopian SF novel *Brave New World* (1932).

Rental: Director Ken Russell's *The Devils* (1971) was based on the Huxley book. It delves into the political machinations behind the trial, as well as the epistemological problems.

Descartes' evil demon hypothesis is the statement that *it is possible that there is an evil demon that systematically deceives me about all of my beliefs.* Here is one way to reframe this hypothesis and its accompanying thought experiment as a skeptical argument.

(2.15) I know any statement S only if I know that the evil demon hypothesis is false.

(2.16) I cannot know that the evil demon hypothesis is false.

(2.17) Therefore, I do not know any statement S, in other words, I know nothing.

The truth of premise 2.15 depends upon Descartes' definition of the key term "knowledge" as requiring indubitable justification.

2.16 is justified on the grounds that there is no first-person criterion for determining whether or not I am being deceived by an evil demon. Any evidence to which I appeal to disprove the evil demon hypothesis will fail. Arguing that I have no reason to believe that an evil demon actually exists does *not* disprove the evil demon hypothesis. True, I don't have reason to believe that any evil demon exists. But more importantly, I don't have good reasons to believe that it is *not possible* that an evil demon exists. The evil demon hypothesis only appeals to the *possible* existence of an evil demon. If I cannot disprove the evil demon hypothesis, then I am unable to prove that I know anything.

We can appreciate Descartes' skeptical arguments by a comparison with science fiction scenarios. One interesting debate regards whether skeptical science fiction scenarios parallel dream skepticism more than they do **evil demon skepticism**. When describing *The Matrix* just now, we wrote that the humans in the film seem to themselves to live the humdrum sorts of lives that we on Earth actually do live. How do I know, given the conceivability of such developments, that I am not deceived just as the characters in such stories are deceived?

One necessary condition for evil demon skepticism is that agents are unable to determine the difference between appearance and reality. I know the difference between dreaming and waking experience, but I simply don't know, at a given point in time, which I am having at that moment. We just observed that in *The Matrix*, the main characters not only can know the difference but in fact do know the difference between real and virtual experience. The same is eventually true of Captain Picard in "Ship in a Bottle."

The film *Existenz* stands apart from *The Matrix* and *Star Trek* because only *Existenz* makes the point that a virtual reality environment can itself contain other virtual environments. (*Rental*: *Existenz* (1999) directed by David Cronenberg and starring Jude Law and Jennifer Jason Leigh.) This makes *Existenz* epistemically challenging in ways that *The Matrix* films are not. In *Existenz*, candidates are selected for a special beta-test of a new virtual reality game. They assemble and begin playing in one virtual world, only to be shifted from one world to the other, never knowing which amongst those worlds is the real world. But once in the virtual environment they cannot determine which of countless "worlds" is real. In other words, *Existenz* illustrates strong evil demon skepticism whereas *The Matrix* only illustrates a weaker form of dream skepticism.

2.4 ASSESSING THE SKEPTICAL ARGUMENTS

Consider again the argument from dreaming. The first premise states that there is some qualitative similarity between what seems to be dream experience and what seems to be waking experience. But perhaps I have a test for determining whether I'm dreaming: in dreams I do not feel pain. So I can pinch myself and determine whether I'm dreaming based on whether I feel pain as a result. This shows that the claim about qualitative similarity in premise 2.8 is false.

But to this the skeptic has a ready response. For any test that I consider, it is possible to *dream* that I perform the test. And, furthermore, I can dream any result to any such test. Therefore, I can't rule out the possibility that, at any given moment, I am dreaming. This objection to premise 2.8 begs the question and presupposes that premise 2.8 is false. The critic claims that in dreams I do not experience pain—but this is only to assume that I know when I am dreaming and when I am not!

The same sort of point can be made in response to a slightly different qualitative test. A critic might observe that in what seems to be waking experience, there is a greater degree of coherence and regularity than there is in what seems to be dream experience. For examples, in what appears to be waking life, laws of nature are not violated and I don't find myself performing actions wholly out of my character. But again, this type of objection only presupposes that we know the difference between dreaming and waking experience. Perhaps, the skeptic retorts, it is my dreaming experience that is the more coherent and regular. This points to the insidiousness of skeptical arguments; they are intellectual quicksand.

A famous response to global skepticism was offered by Descartes himself, in the second of his *Meditations*. It begins with a question: are there any beliefs whatsoever about which I can be certain? Suppose I am in a virtual world, wherein the intelligence that deceives me can even make false my belief that $1 + 1 = 2$. Is there anything at all that I can know with certainty?

Descartes believes that I know one statement with certainty: that I exist. (We can see why Descartes chooses to write in the first-person.) Even if I might be deceived in all I think,

says Descartes, still I am certain of this: I think, I am. For even if *what* I think is false and deceptive, *that* I think cannot be doubted, for to doubt is to think, and, even if my thoughts are deceived, they are still thoughts. If I think, then I am, I exist. This thought is called the **cogito**, from the Latin "I think." It rests upon the fact that statements such as this are *self-verifying*: by doubting the statement in question, by doubting that I think, I thereby think. In other words, since doubting is a form of thinking, I make the statement true by doubting it.

Establishing this belief as true and sufficiently justified marks the first half of Descartes' move out of global skepticism. From this foundation, as small as it may seem, Descartes proceeds to argue that we can once again come to know most of the beliefs that he initially subjected to doubt. Descartes believes that from this modest beginning he can offer arguments that give us certain knowledge of the existence of God, and of the physical world.

But there are other interesting ways to respond to evil demon skepticism. First, the premises of the argument depend upon a privileged definition of "know": for Descartes, to know a statement entails that the statement is indubitable for you. You cannot doubt it. If we simply lower those demanding standards for knowledge, we can deny his claims that we do not "know" certain statements. For example, in order to know something perhaps I only need to have pretty good evidence on its behalf. If so, then my knowledge of beliefs from memory, perception or mathematics are known *even though* I can't show that it is impossible an evil demon is deceiving me. (We explore this response to skepticism shortly in section §2.9.)

BOX 2.G: DID DESCARTES PLAGIARIZE ST. AUGUSTINE?

It wasn't common practice to cite sources in the 1640s as it is today. But we shouldn't forget that Saint Augustine (b. 354–d. 430 AD) raised the skeptical worries Descartes did, and solved them in the same ways too. Saint Augustine writes,

> In regard to this [being deceived about our knowledge that we live] we are absolutely without any fear unless we are perhaps being deceived by some resemblance of the truth; since it is certain, that he who is deceived, yet lives . . . The knowledge by which we know that we live is the most inward of all knowledge, of which even the Academic [skeptic] cannot insinuate: Perhaps you are asleep, and do not know it . . . Nor can the academic again say, in confutation of this knowledge: Perhaps you are mad, and do not know it: . . . but he who is mad is alive. Therefore he who says he knows he is alive, can neither be deceived nor lie. Let a thousand kinds of deceitful objects of sight

> be presented to him who says, I know I am alive; yet he will fear none of them, for he who is deceived yet is alive.
>
> (*De Trinitate* XV.xii.21)

Not only does Augustine concisely explain the epistemological status of self-verifying propositions, but he also seems to take up just those skeptical considerations Descartes later does: about bodily senses, sleep and dreaming, and insanity.

The refutation of global skepticism from the indubitability of self-verifying propositions like "I think" is most clear in Augustine's most famous work, *The City of God*:

> In the face of these truths, the quibbles of the skeptics lose their force. If they say, "What if you are mistaken?"—well, if I am mistaken, I am. For, if one does not exist, he can by no means be mistaken. Therefore, I am, if I am mistaken. Because, therefore, I am, if I am mistaken, how can I be mistaken that I am, since it is certain that I am, if I am mistaken? and because, if I could be mistaken, I would have to be the one who is mistaken, therefore, I am most certainly not mistaken in knowing that I am. Nor, as a consequence, am I mistaken in knowing that I know. For, just as I know that I am, I also know that I know.
>
> (*The City of God*, book XI, c.26)

Augustine rarely receives the credit he deserves for foreseeing this "Cartesian" solution to global skepticism. Not only has he foreseen the epistemological utility of self-verifying statements, he seems to be better attuned to the subtlety of such statements than Descartes himself. You can be the judge as to whether Descartes is guilty of plagiarism. Perhaps a better question is: Did a demon make him do it?

2.5 FOUNDATIONALISM

The structure of Descartes' replies to his own skeptical arguments influences contemporary epistemology. Because he starts from a few fundamental truths and builds from these a broad structure of knowledge, Descartes is called a foundationalist. **Foundationalism** is the thesis that special statements called *basic beliefs* are known and that basic beliefs allow us to justify further beliefs based upon them. What is special about basic beliefs is that they are known but their knowledge does not depend upon the knowledge of any other beliefs. In geometry, one starts with a few basic beliefs called "axioms," and one uses these axioms to construct theorems. Using the axioms, one can generate a very sophisticated system of thought. The foundationalist thinks that a similar structure can be constructed for ordinary knowledge claims, on the basis of a few basic beliefs.

The central motivation for foundationalism involves the **regress of knowledge** problem. My belief B is justified and true, and so is knowledge. The justification for B is another belief P. P justifies B for me because P is also justified and true: I know P too. But we have a similar pattern with respect to the justification of P. P is justified by my belief that Q, which is in turn justified and true. But how do I justify Q, and the beliefs upon which Q is itself justified? This sequence of beliefs can take one of only a few forms. It may (i) loop back upon itself; (ii) may continue on indefinitely; (iii) stop in beliefs that are themselves unknown; or (iv) it may stop in basic beliefs that are known independently of knowledge of yet other beliefs.

The most plausible responses to the regress of knowledge problem are (i) and (iv). *Coherentists* are those philosophers who adopt a form of option (i). Foundationalists opt for (iv). Few philosophers are attracted to options (ii) and (iii) because it is thought that neither of those options achieves the desired result—showing that I know B.

Coherentists opt for a form of (i), but this option can be given different logical structures. The loop might be a linear circle: Q justifies P, which justifies B, and B justifies Q, which . . . This chain-like structure does not seem to justify all its component beliefs. If that were the structure of my justification for B, the justification would be inadequate to convert belief B into knowledge. But coherentists argue for a version of (i) in which the structure of our beliefs resembles a spider's web. In a spider's web each filament is connected to every other filament through very strong ties. The overall result is a stable, well engineered structure. The coherentist holds that our network of beliefs is similarly interconnected and produces a stable structure. Member beliefs bear numerous inferential relationships with other beliefs, producing a coherent nexus of beliefs to support knowledge.

Foundationalists argue, in contrast, that option (iv) best solves the regress problem. Foundationalists are led to find basic beliefs that are known, but whose knowledge does not depend on the knowledge of other beliefs. Basic beliefs serve as the foundation for the superstructure of beliefs that we build on top of them. When we construct a big, multi-storey building, the only way it can be made to stand is by pouring a solid foundation. Reflect on the enormously layered nature of your belief system. The foundationalist suggests that beliefs in those upper layers—layers containing no basic beliefs at all—crucially depend for their justification upon lower layers that are well grounded and well justified.

If a skeptic shows that foundational beliefs are uncertain, he effectively shows that all beliefs in the structure are uncertain. Such a skeptic might argue in this way against Descartes by attempting to raise doubts about whether Descartes knows with certainty that he exists.

BOX 2.H: PUTNAM ON RESOLVING BRAIN-IN-A-VAT SKEPTICISM

A contemporary philosopher, Hilary Putnam, describes what he calls *brain-in-a-vat skepticism* and articulates a novel response to it. His response uses resources from

the philosophy of language to make a point about the limits of the ability to refer to objects with words.

> Imagine that a human being (you can imagine this to be yourself) has been subjected to an operation by an evil scientist. The person's brain (your brain) has been removed from the body and placed in a vat of nutrients which keeps the brain alive. The nerve endings have been connected to a super-scientific computer which causes the person whose brain it is to have the illusion that everything is perfectly normal. There seem to be people, objects, the sky, etc., but really all the person (you) is experiencing is the result of electronic impulses traveling from the computer to the nerve endings. The computer is so clever that if the person tries to raise his hand, the feedback from the computer will cause him to "see" and "feel" the hand being raised. Moreover, by varying the program, the evil scientist can cause the victim to "experience" (or hallucinate) any situation or environment the evil scientist wishes. He can also obliterate the memory of the brain operation, so that the victim will seem to himself to have always been in this environment. It can even seem to the victim that he is sitting and reading these very words about the amusing but quite absurd supposition that there is an evil scientist who removes people's brains from their bodies and places them in a vat of nutrients which keep the brains alive.
>
> (Putnam, 1981: pp. 5–6)

Putnam argues that he can refute brain-in-a-vat skepticism on grounds having to do with the nature of language.

Suppose that the meanings of the words used by the envatted person are fixed by their causal connections with objects in the mind-independent world, as opposed to being fixed by networks of mental states in the mind. So his thoughts about fire hydrants have the structure and content that they do in part because he has experienced causal interactions with fire hydrants. If so, then, if he were a brain-in-a-vat, it would be impossible for him to form the thought that he is a brain-in-a-vat since that thought fails to bear the appropriate causal connections to the mind-independent world! When a brain-in-a-vat thinks "that's a fire hydrant," the term "fire hydrant" means something different in brain-in-a-vat-English than it does in normal English. This argument is based upon a theory in the philosophy of language known as *semantic externalism*.

There are a number of different proposals on how to use semantic externalism to prove that brain-in-a-vat skepticism is false. The most commonsensical one is this: (i) Brains in vats fail to have concepts through which they can think about things like cars, Mars and candy bars; (ii) I, though, have just the sorts of concepts needed to think about those things; (iii) therefore, I know I am not a brain-in-a-vat. Concerns about this use of semantic externalism to refute skepticism have focused upon (ii), for just how the non-skeptic justifies this premise without begging the question against the skeptic is unclear.

Despite its apparent decisiveness in refuting skepticism, there are actually several questions we can raise about the *cogito*. Foremost among these is: What is its precise logical status? Many have thought that the *cogito* and other propositions like it are *self-verifying*, that is, known in virtue of being believed. But is the *cogito* self-verifying? Descartes states the *cogito* in two different ways. In the *Meditations*, he puts it as we have above: "I think, I am." But in a lesser known book, *The Rules for the Direction of Mind*, he further writes, "I think, therefore I am" (in Latin *Cogito, ergo sum*). The way of putting the point in *The Rules* makes it seem as though the *cogito* is an inference: "I think" is the premise, and "I am" is the conclusion. If interpreted in this way, though, "I think, so I exist" begs the question since my existence, the conclusion, is presupposed in the premise.

Yet another way in which a skeptic might argue against Descartes' foundationalist approach is to say that, even if the foundation is indubitable and certain, it is inadequate to support anything beyond itself. The foundationalist needs to have basic knowledge of statements that are both certain and sufficiently numerous to be able to use them to infer further truths. If all I can rightly claim to know is that I think and exist, I am not far from being a global skeptic. The global skeptic will ask, "When you make inferences from knowing that you exist to knowing statements about what your dog ate yesterday, how do you know that your reasoning isn't confused by the manipulations of the evil demon?" The content of my belief in my existence doesn't lend any evidential support to my belief about my dog—they are about wholly different things. Descartes must find a set of basic beliefs from which he can build an epistemological bridge from what he knows to what he does not yet know. No bridge between these two beliefs is apparent. So it seems Descartes may need more beliefs in his foundations for knowledge. But "I think" and other beliefs very similar to it are the only beliefs capable of meeting Descartes' indubitability requirement for knowledge.

2.6 CLOSURE PRINCIPLES

The skeptic insists that I can't know that skeptical hypotheses, like the evil demon hypothesis, are false. We see this in premise 2.15: I know some statement P only if I know that the skeptical hypothesis is false. This wouldn't be a problem if it weren't for the fact that, as stated in 2.16, I cannot know that the skeptical hypothesis is false. It is the combination of 2.15 with 2.16 that justifies the conclusion that I have no knowledge. Implicitly, these two premises are bound logically together by a conditional statement of the form:

(2.18) If I do not know that skeptical hypotheses are false, then I do not know most of the beliefs I think I know.

Typically, the first line of attack on skeptical arguments is to argue against premises like 2.15 by giving specific refutations of each type of skeptical hypothesis. This is what we have been engaged in doing (so far, unsuccessfully) when assessing forms of skepticism. But there is another important second line of attack open to the non-skeptic in which he targets the logical relationship represented in 2.18.

This line of attack uses **closure principles** about knowledge. *Closure principles* specify the conditions under which one is justified to move from some known belief P to knowledge of another belief Q. They can also specify conditions under which one is not justified to move from some known belief P to knowledge of another belief Q.

The non-skeptic may turn the tables on the skeptic's reasoning in 2.15 to 2.18 by using a closure principle like this:

(2.19) If an agent knows P, and knows that if P then Q, then the agent knows Q.

This principle can be used in a refutation of skepticism. Consider the following two statements:

(P) I see that snow is falling in New York City.

(Q) I am not now dreaming.

Suppose I know P because I am in New York City, and I see snow falling all around me. If P—that is, if I am actually seeing snow falling in New York City—then I cannot be dreaming. In dreams, I do not actually see anything and I am not actually in New York City. In dreams I only *seem* to see and *appear* to be somewhere else. In this case, if 2.19, P and Q combine to justify the conclusion that dream skepticism is false.

When we attempt to assess whether an agent knows a proposition, the agent is thought to need to rule out alternative beliefs that, if true, would defeat the agent's knowledge of the proposition. Consider an alternative to P, namely:

(S) All my perceptual beliefs, including the belief that it is snowing in New York City, are the products of a virtual reality simulation I am now running.

If the context in which P is asserted is a local radio station's weather report, then the skeptical possibility presented in S is not relevant. Alternatives are deemed relevant or irrelevant in part upon the context in which the statement is asserted. But if the context in which P is asserted is a complex science fiction story involving artificial intelligences that harbor prejudices against humanity, then S becomes relevant.

On the basis of this use of the non-skeptic's closure principle, non-skeptics argue that in order to know most of the beliefs I think I know, I do not need to know that skeptical hypotheses are false because such skeptical hypotheses are irrelevant to my knowledge of most of what I know. Therefore, the non-skeptic alleges, the problem of global skepticism is overcome.

But the skeptic has a reply at hand. In the context of a radio station's weather report, other alternatives are relevant. Consider this one:

(R) I see that fine flakes of white ash emitted by a factory are falling in New York City.

Let's set the context. It is a balmy July in New York, where the daytime temperature hasn't dropped below 85 for weeks. It happens that there was an industrial accident in the city that morning that has emitted 30 tonnes of white ash into the sky. One radio station's local weather watchers report to the station that it is snowing. In addition, many background beliefs play important epistemic roles in making a judgment about R. These include the beliefs that snow only falls when it is cold outside, for example.

With the context and background assumptions in mind, R is relevant given the context. The local weather watcher's report is not justified because he has failed to rule out R as an alternative explanation of his sensory perception. So there are circumstances in which some alternatives are relevant, and serve to scuttle the knowledge that one claims to possess. The skeptic directs us to contexts in which an alternative like S is relevant to my belief that P. What would such a context be? One strong candidate for a context in which S is a relevant alternative to P is the very context we're in now—doing philosophy. The skeptic will argue that the use of 2.19 makes a travesty of the problems of epistemology. Skepticism is refuted in one exceedingly simple step if the non-skeptic's closure principle is true. But, the skeptic argues, this refutation is much too easy. Invoking the non-skeptic's closure principle is tantamount to dissolving the skeptical challenge, and does not take the problem seriously. The non-skeptic wins the bet only by fixing the race, or so epistemologists have argued.

2.7 CERTAINTY AND FALLIBILISM

Perhaps, after considering skeptical arguments, you thought that they were fanciful and unrealistic, that they took the problem of knowledge too seriously in the first place. If so, you are in good company. *Fallibilists* in epistemology attempt to restructure our thinking about knowledge by arguing that person A can know statement P even if A's justification for P is not certain or indubitable. Fallibilists argue that Descartes took epistemology down a dead-end with his exceedingly strong requirements on knowledge. Most knowledge claims do not meet the requirements set out by Descartes. But fallibilists take this as a reason to reject those requirements in the first place.

Fallibists differ from skeptics on an important point. Skeptics claim that beliefs fail to meet a high standard for knowledge, and they conclude that we lack knowledge. Fallibilists claim that beliefs fail to meet a high standard for knowledge, but they do not conclude that we have no knowledge. Fallibilists just endorse different standards for knowledge. Typically, if someone were to ask you, "Do you know that you are reading a book?," you would answer, "Of course!" But then he asks, "Couldn't it be the case that an evil demon was deceiving you into thinking that you are reading a book, when in fact you are not?" Now you might answer, "Well, it's possible that I am mistaken in such a way, but I never claimed to be *infallible*. Nonetheless, I still believe that I know I am reading a book." If you answer in this way, you are a fallibilist. According to **fallibilism**, you can have knowledge of P even if P is not certain for you. The fallibilist is then able to argue that we can know most of what we think we know.

But, as might be expected, many are not convinced that this answer defeats skepticism. We can imagine a case in which the majority of people used a term incorrectly. For example, the proper definition of "Martian" might be "an intelligent creature from Mars." Now that we have landed very sophisticated equipment on Mars and have found it to be devoid of life, we might conclude that there are no Martians. But then someone might come along and say that most people use the term "Martian" to refer to any alien from another world, as a synonym for "extraterrestrial being." Does that mean that when we say, "There are no Martians because there is no life on Mars," we have employed a poor analysis of the term "Martian?" No, it only means that, in ordinary usage, people use the term "Martian" carelessly.

Similarly, the skeptic might argue that the fact that people use the word "knowledge" in all sorts of careless ways does not show that the analysis of knowledge used by the skeptic is incorrect. It may only show that a careless use of the term is common. The skeptic would conclude that, on a proper understanding of the term, it refers to few (if any) actual cases. If we choose to use the word "knowledge" in a fallibilist way, we are simply using another concept. We are not refuting the skeptic's arguments.

The skeptic challenges us to defend our claim that we have a large amount of knowledge about various subjects (or, in the case of global, evil demon skepticism, about anything at all). The skeptic's strategy is to point out that many of the beliefs we have that we think are knowledge do not qualify as knowledge. Many of our beliefs about scientific matters are false, for example, but I rely equally on them all and count them all as knowledge. We must grant to the skeptic that at least some of what we currently count as knowledge is not. Then the question becomes: *Which* of our beliefs are known?

2.8 SUMMARY

Chapter 2 began with a discussion of the concept of knowledge and a definition of that term. Propositional knowledge is our focus, and we offered a justified-true-belief definition of it. This, however, was subject to the Agatha counterexample. Her belief *that it is 10 a.m.* is justified and true, but it is insufficient for knowledge since its truth depends upon a lucky accident.

We examined dream and evil demon skepticisms, which Descartes uses in his *Meditations*. We presented and evaluated arguments on behalf of those forms of skepticism. Criticisms of those arguments included appeals to the "cogito," to foundationalism and to closure principles.

According to foundationalism there are basic beliefs, and they are self-justifying. The regress of justification—the demand to justify all those beliefs that justify other beliefs—stops when it reaches basic beliefs., so says the foundationalist. Descartes himself was a foundationalist whose only foundational principle was the cogito. That has limited success in stopping skepticism, however, because that foundation is ineffective at justifying a wide selection of other beliefs.

This led us into a discussion of closure principles. Closure principles about knowledge specify the conditions under which one is justified in inferring an unknown belief from a known belief. Both skeptics and non-skeptics use these principles to debate the burden of proof. We concluded the chapter by discussing the role of certainty in refutations of skepticism. Descartes sets the bar for knowledge very high indeed; philosophers since his time have adopted weaker definitions of knowledge and fallible foundations in their continuing engagement with skepticism.

2.9 ABOUT THE READINGS

This chapter includes four readings, one of science fiction, one from ancient philosophy, one from early modern philosophy and a final contemporary philosophy paper.

Our science fiction selection, "We Can Remember It for You Wholesale," is perhaps the best Philip K. Dick exploration of the boundaries between appearance and reality. Like other Dick stories, this was produced as a film under the title *Total Recall* (1990). Douglas Quail is a miserable salaried employee for the West Coast Emigration Bureau, but he has one abiding dream—to visit Mars before he dies. He can't afford the trip, so he settles for implanted memories of such a trip, which he buys at Rekal Inc. But after the doctors at Rekal begin the procedure, they discover that Quail has a number of apparently false memories embedded in his brain already. They hustle him out the door with a refund, but procedure has brought back to Quail fragmentary memories of a Mars trip. Which are real and which are fake? Neither he nor Rekal seem to know. Quail learns the truth when police from the security agency Interplan hunt him down. Or does he?

Plato, in *Republic* book VII, explores the appearance/reality with a brief story of his own. In the "Allegory of the Cave," Plato offers compelling visual imagery to describe the inability of the human mind to gain knowledge of reality as it is in itself. He believed that the physical world was illusory. Knowledge, he said, must be of things that do not change. But we are surrounded by a physical environment that changes every moment. We are like prisoners who see only shadows on the wall. And shadows of what? Puppets. We are benighted and deceived.

This allegory can be compared and contrasted with Descartes' skepticism. Descartes also believed that our senses do not give us knowledge of the world and, like Plato, believed that reason was infinitely better than the senses at acquiring knowledge. In the 1640s, when the *Meditations* were published, questioning the prejudices and presuppositions of the authorities was punishable by death. For example, shortly before Descartes' career as an author, Galileo was brought a second time before an inquisition by the Catholic Church on the grounds that he believed in a heliocentric solar system, which contradicted the Church's teachings and conflicted with the Bible. With stakes so high, Descartes adopted a literary form that allowed him some freedom to speak his mind in order to achieve his goal of laying waste to the inherited beliefs passed down by the Catholic Church of his day. This philosophical treatise was put in the form of six daily meditations, and the selection

here includes the first and part of the second. There he presents the case for dream skepticism, evil demon skepticism and his response to those arguments with the claim that "I think, I exist."

Contemporary philosophers have found ways to amplify the force of skepticism by appeal to the power of computer simulations. The simulation argument, developed by Nick Bostrom of the Future of Humanity Institute at Oxford University, contends that one of the following claims must be true: (1) intelligent beings won't reach a level of technology at which they create simulations so sophisticated that they are indistinguishable from reality; (2) species that do reach such a level refrain from creating them; or (3) it is highly probable that we are living in a simulation, not in the real world. In our reading, Alasdair Richmond, of Edinburgh University, analyzes the simulation argument, compares its content with other skeptical arguments in the history of philosophy (including Descartes'), and diagnoses some problems with it.

2.10 "WE CAN REMEMBER IT FOR YOU WHOLESALE," BY PHILIP K. DICK

He awoke—and wanted Mars. The valleys, he thought. What would it be like to trudge among them? Great and greater yet: the dream grew as he became fully conscious, the dream and the yearning. He could almost feel the enveloping presence of the other world, which only Government agents and high officials had seen. A clerk like himself? Not likely.

"Are you getting up or not?" his wife Kirsten asked drowsily, with her usual hint of fierce crossness. "If you are, push the hot coffee button on the darn stove."

"Okay," Douglas Quail said, and made his way barefoot from the bedroom of their conapt to the kitchen. There, having dutifully pressed the hot coffee button, he seated himself at the kitchen table, brought out a yellow, small tin of fine Dean Swift snuff. He inhaled briskly, and the Beau Nash mixture stung his nose, burned the roof of his mouth. But still he inhaled; it woke him up and allowed his dreams, his nocturnal desires and random wishes, to condense into a semblance of rationality.

I will go, he said to himself. *Before I die I'll see Mars*.

It was, of course, impossible, and he knew this even as he dreamed. But the daylight, the mundane noise of his wife now brushing her hair before the bedroom mirror—everything conspired to remind him of what he was. *A miserable little salaried employee*, he said to himself with bitterness. Kirsten reminded him of this at least once a day and he did not blame her; it was a wife's job to bring her husband down to Earth. *Down to Earth*, he thought, and laughed. The figure of speech in this was literally apt.

"What are you sniggering about?" his wife asked as she swept into the kitchen, her long busy-pink robe wagging after her. "A dream, I bet. You're always full of them."

"Yes," he said, and gazed out the kitchen window at the hover-cars and traffic runnels, and all the little energetic people hurrying to work. In a little while he would be among them. As always.

"I'll bet it has to do with some woman," Kirsten said witheringly.

"No," he said. "A god. The god of war. He has wonderful craters with every kind of plant-life growing deep down in them."

"Listen." Kirsten crouched down beside him and spoke earnestly, the harsh quality momentarily gone from her voice. "The bottom of the ocean—*our* ocean is much more, an infinity of times more beautiful. You know that; everyone knows that. Rent an artificial gill-outfit for both of us, take a week off from work, and we can descend and live down there at one of those year-round aquatic resorts. And in addition," She broke off. "You're not listening. You should be. Here is something a lot better than that compulsion, that obsession you have about Mars, and you don't even listen!" Her voice rose piercingly. "God in heaven, you're doomed, Doug! What's going to become of you?"

"I'm going to work," he said, rising to his feet, his breakfast forgotten. "That's what's going to become of me."

She eyed him. "You're getting worse. More fanatical every day. Where's it going to lead?"

"To Mars," he said, and opened the door to the closet to get down a fresh shirt to wear to work.

Having descended from the taxi Douglas Quail slowly walked across three densely-populated foot runnels and to the modern, attractively inviting doorway. There he halted, impeding mid-morning traffic, and with caution read the shifting-color neon sign. He had, in the past, scrutinized this sign before . . . but never had he come so close. This was very different; what he did now was something else. Something which sooner or later had to happen.

Rekal, Incorporated

Was this the answer? After all, an illusion, no matter how convincing, remained nothing more than an illusion. At least objectively. But subjectively—quite the opposite entirely.

And anyhow he had an appointment. Within the next five minutes. Taking a deep breath of mildly smog-infested Chicago air, he walked through the dazzling polychromatic shimmer of the doorway and up to the receptionist's counter. The nicely-articulated blonde at the counter, bare-bosomed and tidy, said pleasantly, "Good morning, Mr. Quail."

"Yes," he said. "I'm here to see about a Rekal course. As I guess you know."

"Not "rekal' but *re*call," the receptionist corrected him. She picked up the receiver of the vidphone by her smooth elbow and said into it, "Mr. Douglas Quail is here, Mr. McClane. May he come inside, now? Or is it too soon?"

"Giz wetwa wum-wum wamp," the phone mumbled.

"Yes, Mr. Quail," she said. "You may go on in; Mr. McClane is expecting you." As he started off uncertainly she called after him, "Room D, Mr. Quail. To your right."

After a frustrating but brief moment of being lost he found the proper room. The door hung open and inside, at a big genuine walnut desk, sat a genial-looking man, middle-aged, wearing the latest Martian frog-pelt gray suit; his attire alone would have told Quail that he had come to the right person.

"Sit down, Douglas," McClane said, waving his plump hand toward a chair which faced the desk. "So you want to have gone to Mars. Very good."

Quail seated himself, feeling tense. "I'm not so sure this is worth the fee," he said. "It costs a lot and as far as I can see I really get nothing." *Costs almost as much as going*, he thought.

"You get tangible proof of your trip," McClane disagreed emphatically. "All the proof you'll need. Here; I'll show you." He dug within a drawer of his impressive desk. "Ticket stub." Reaching into a manila folder he produced a small square of embossed cardboard. "It proves you went—and returned. Postcards." He laid out four franked picture 3-D full-color postcards in a neatly-arranged row on the desk for Quail to see. "Film. Shots you took of local sights on Mars with a rented movie camera." To Quail he displayed those, too. "Plus the names of people you met, two hundred poscreds worth of souvenirs, which will arrive—from Mars—within the following month. And passport, certificates listing the shots you received. And more." He glanced up keenly at Quail. "You'll know you went, all right," he said. "You won't remember us, won't remember me or ever having been here. It'll be a real trip in your mind; we guarantee that. A full two weeks of recall; every last piddling detail. Remember this: if at any time you doubt that you really took an extensive trip to Mars you can return here and get a full refund. You see?"

"But I didn't go," Quail said. "I won't have gone, no matter what proofs you provide me with." He took a deep, unsteady breath. "And I never was a secret agent with Interplan." It seemed impossible to him that Rekal, Incorporated's extra-factual memory implant would do its job—despite what he had heard people say.

"Mr. Quail," McClane said patiently. "As you explained in your letter to us, you have no chance, no possibility in the slightest, of ever actually getting to Mars; you can't afford it, and what is much more important, you could never qualify as an undercover agent for Interplan or anybody else. This is the only way you can achieve your, ahem, life-long dream; am I not correct, sir? You can't be this; you can't actually do this." He chuckled. "But you can *have been* and *have done*. We see to that. And our fee is reasonable; no hidden charges." He smiled encouragingly.

"Is an extra-factual memory that convincing?" Quail asked.

"More than the real thing, sir. Had you really gone to Mars as an Interplan agent, you would by now have forgotten a great deal; our analysis of true-mem systems—authentic recollections of major events in a person's life—shows that a variety of details are very quickly lost to the person. Forever. Part of the package we offer you is such deep implantation of recall that nothing is forgotten. The packet which is fed to you while you're comatose is the creation of trained experts, men who have spent years on Mars; in every case we verify details down to the last iota. And you've picked a rather easy extra-factual system; had you picked Pluto or wanted to be Emperor of the Inner Planet Alliance we'd have much more difficulty . . . and the charges would be considerably greater."

Reaching into his coat for his wallet, Quail said, "Okay. It's been my life-long ambition and I can see I'll never really do it. So I guess I'll have to settle for this."

"Don't think of it that way," McClane said severely. "You're not accepting second-best. The actual memory, with all its vagueness, omissions and ellipses, not to say distortions—that's second-best." He accepted the money and pressed a button on his desk.

"All right, Mr. Quail," he said, as the door of his office opened and two burly men swiftly entered. "You're on your way to Mars as a secret agent." He rose, came over to shake Quail's nervous, moist hand. "Or rather, you have been on your way. This afternoon at four-thirty you will, um, arrive back here on Terra; a cab will leave you off at your conapt and as I say you will never remember seeing me or coming here; you won't, in fact, even remember having heard of our existence."

His mouth dry with nervousness, Quail followed the two technicians from the office; what happened next depended on them.

Will I actually believe I've been on Mars? he wondered. *That I managed to fulfill my lifetime ambition?* He had a strange, lingering intuition that something would go wrong. But just what—he did not know.

He would have to wait to find out.

The intercom on McClane's desk, which connected him with the work area of the firm, buzzed and a voice said, "Mr. Quail is under sedation now, sir. Do you want to supervise this one, or shall we go ahead?"

"It's routine," McClane observed. "You may go ahead, Lowe; I don't think you'll run into any trouble." Programming an artificial memory of a trip to another planet—with or without the added fillip of being a secret agent—showed up on the firm's work-schedule with monotonous regularity. *In one month*, he calculated wryly, *we must do twenty of these . . . ersatz interplanetary travel has become our bread and butter.*

"Whatever you say, Mr. McClane," Lowe's voice came, and thereupon the intercom shut off.

Going to the vault section in the chamber behind his office, McClane searched about for a Three packet—trip to Mars—and a Sixty-two packet: secret Interplan spy. Finding the two packets, he returned with them to his desk, seated himself comfortably, poured out the contents—merchandise which would be planted in Quail's conapt while the lab technicians busied themselves installing the false memory.

A one-poscred sneaky-pete side arm, McClane reflected; *that's the largest item. Sets us back financially the most.* Then a pellet-sized transmitter, which could be swallowed if the agent were caught. Code book that astonishingly resembled the real thing . . . the firm's models were highly accurate: based, whenever possible, on actual U.S. military issue. Odd bits which made no intrinsic sense but which would be woven into the warp and woof of Quail's imaginary trip, would coincide with his memory: half an ancient silver fifty cent piece, several quotations from John Donne's sermons written incorrectly, each on a separate piece of transparent tissue-thin paper, several match folders from bars on Mars, a stainless steel spoon engraved PROPERTY OF DOME-MARS NATIONAL KIBBUZIM, a wire tapping coil which—

The intercom buzzed. "Mr. McClane, I'm sorry to bother you but something rather ominous has come up. Maybe it would be better if you were in here after all. Quail is already under sedation; he reacted well to the narkidrine; he's completely unconscious and receptive. But—"

"I'll be in." Sensing trouble, McClane left his office; a moment later he emerged in the work area.

On a hygienic bed lay Douglas Quail, breathing slowly and regularly, his eyes virtually shut; he seemed dimly—but only dimly—aware of the two technicians and now McClane himself.

"There's no space to insert false memory-patterns?" McClane felt irritation. "Merely drop out two work weeks; he's employed as a clerk at the West Coast Emigration Bureau, which is a government agency, so he undoubtedly has or had two weeks vacation within the last year. That ought to do it." Petty details annoyed him. And always would.

"Our problem," Lowe said sharply, "is something quite different." He bent over the bed, said to Quail, "Tell Mr. McClane what you told us." To McClane he said, "Listen closely."

The gray-green eyes of the man lying supine in the bed focused on McClane's face. The eyes, he observed uneasily, had become hard; they had a polished, inorganic quality, like semi-precious tumbled stones. He was not sure that he liked what he saw; the brilliance was too cold. "What do you want now?" Quail said harshly. "You've broken my cover. Get out of here before I take you all apart." He studied McClane. "Especially you," he continued. "You're in charge of this counter-operation."

Lowe said, "How long were you on Mars?"

"One month," Quail said gratingly.

"And your purpose there?" Lowe demanded.

The meager lips twisted; Quail eyed him and did not speak. At last, drawling the words out so that they dripped with hostility, he said, "Agent for Interplan. As I already told you. Don't you record everything that's said? Play your vid-aud tape back for your boss and leave me alone." He shut his eyes, then; the hard brilliance ceased. McClane felt, instantly, a rushing splurge of relief.

Lowe said quietly, "This is a tough man, Mr. McClane."

"He won't be," McClane said, "after we arrange for him to lose his memory-chain again. He'll be as meek as before." To Quail he said, "So *this* is why you wanted to go to Mars so terribly bad."

Without opening his eyes Quail said, "I never wanted to go to Mars. I was assigned it—they handed it to me and there I was: stuck. Oh yeah, I admit I was curious about it; who wouldn't be?" Again he opened his eyes and surveyed the three of them, McClane in particular. "Quite a truth drug you've got here; it brought up things I had absolutely no memory of." He pondered. "I wonder about Kirsten," he said, half to himself. "Could she be in on it? An Interplan contact keeping an eye on me . . . to be certain I didn't regain my memory? No wonder she's been so derisive about my wanting to go there." Faintly, he smiled; the smile—one of understanding—disappeared almost at once.

McClane said, "Please believe me, Mr. Quail; we stumbled onto this entirely by accident. In the work we do—"

"I believe you," Quail said. He seemed tired, now; the drug was continuing to pull him under, deeper and deeper. "Where did I say I'd been?" he murmured. "Mars? Hard to remember—I know I'd like to see it; so would everybody else. But me—" His voice trailed off. "Just a clerk, a nothing clerk."

Straightening up, Lowe said to his superior, "He wants a false memory implanted that corresponds to a trip he actually took. And a false reason which is the real reason.

He's telling the truth; he's a long way down in the narkidrine. The trip is very vivid in his mind—at least under sedation. But apparently he doesn't recall it otherwise. Someone, probably at a government military-sciences lab, erased his conscious memories; all he knew was that going to Mars meant something special to him, and so did being a secret agent. They couldn't erase that; it's not a memory but a desire, undoubtedly the same one that motivated him to volunteer for the assignment in the first place."

The other technician, Keeler, said to McClane, "What do we do? Graft a false memory-pattern over the real memory? There's no telling what the results would be; he might remember some of the genuine trip, and the confusion might bring on a psychotic interlude. He'd have to hold two opposite premises in his mind simultaneously: that he went to Mars and that he didn't. That he's a genuine agent for Interplan and he's not, that it's spurious. I think we ought to revive him without any false memory implantation and send him out of here; this is hot."

"Agreed," McClane said. A thought came to him. "Can you predict what he'll remember when he comes out of sedation?"

"Impossible to tell," Lowe said. "He probably will have some dim, diffuse memory of his actual trip, now. And he'd probably be in grave doubt as to its validity; he'd probably decide our programming slipped a gear-tooth. And he'd remember coming here; that wouldn't be erased—unless you want it erased."

"The less we mess with this man," McClane said, "the better I like it. This is nothing for us to fool around with; we've been foolish enough to—or unlucky enough to—uncover a genuine Interplan spy who has a cover so perfect that up to now even he didn't know what he was—or rather is." The sooner they washed their hands of the man calling himself Douglas Quail the better.

"Are you going to plant packets Three and Sixty-two in his conapt?" Lowe said.

"No," McClane said. "And we're going to return half his fee."

"Half! Why half?"

McClane said lamely, "It seems to be a good compromise."

As the cab carried him back to his conapt at the residential end of Chicago, Douglas Quail said to himself, *It's sure good to be back on Terra.*

Already the month-long period on Mars had begun to waver in his memory; he had only an image of profound gaping craters, an ever-present ancient erosion of hills, of vitality, of motion itself. A world of dust where little happened, where a good part of the day was spent checking and rechecking one's portable oxygen source. And then the life forms, the unassuming and modest gray-brown cacti and maw-worms.

As a matter of fact he had brought back several moribund examples of Martian fauna; he had smuggled them through customs. After all, they posed no menace; they couldn't survive in Earth's heavy atmosphere.

Reaching into his coat pocket he rummaged for the container of Martian maw-worms—

And found an envelope instead.

Lifting it out he discovered, to his perplexity, that it contained five hundred and seventy poscreds, in 'cred bills of low denomination.

Where'd I get this? he asked himself. *Didn't I spend every 'cred I had on my trip?*

With the money came a slip of paper marked: *one-half fee ret'd. By McClane.* And then the date. Today's date.

"Recall," he said aloud.

"Recall what, sir or madam?" the robot driver of the cab inquired respectfully.

"Do you have a phone book?" Quail demanded.

"Certainly, sir or madam." A slot opened; from it slid a microtape phone book for Cook County.

"It's spelled oddly," Quail said as he leafed through the pages of the yellow section. He felt fear, then; abiding fear. "Here it is," he said. "Take me there, to Rekal, Incorporated. I've changed my mind; I don't want to go home."

"Yes sir, or madam, as the case may be," the driver said. A moment later the cab was zipping back in the opposite direction.

"May I make use of your phone?" he asked.

"Be my guest," the robot driver said. And presented a shiny new emperor 3-D color phone to him.

He dialed his own conapt. And after a pause found himself confronted by a miniature but chillingly realistic image of Kirsten on the small screen. "I've been to Mars," he said to her.

"You're drunk." Her lips writhed scornfully. "Or worse."

"'S God's truth."

"When?" she demanded.

"I don't know." He felt confused. "A simulated trip, I think. By means of one of those artificial or extra-factual or whatever it is memory places. It didn't take."

Kirsten said witheringly, "You *are* drunk." And broke the connection at her end. He hung up, then, feeling his face flush. *Always the same tone*, he said hotly to himself. *Always the retort, as if she knows everything and I know nothing. What a marriage. Keerist*, he thought dismally.

A moment later the cab stopped at the curb before a modern, very attractive little pink building, over which a shifting, polychromatic neon sign read: *Rekal, Incorporated*.

The receptionist, chic and bare from the waist up, started in surprise, then gained masterful control of herself. "Oh hello Mr. Quail," she said nervously. "H-how are you? Did you forget something?"

"The rest of my fee back," he said.

More composed now the receptionist said, "Fee? I think you are mistaken, Mr. Quail. You were here discussing the feasibility of an extra-factual trip for you, but—" She shrugged her smooth pale shoulders. "As I understand it, no trip was taken."

Quail said, "I remember everything, miss. My letter to Rekal, Incorporated, which started this whole business off. I remember my arrival here, my visit with Mr. McClane. Then the two lab technicians taking me in tow and administering a drug to put me out." No wonder the firm had returned half his fee. The false memory of his "trip to Mars" hadn't taken—at least not entirely, not as he had been assured.

"Mr. Quail," the girl said, "although you are a minor clerk you are a good-looking man and it spoils your features to become angry. If it would make you feel any better, I might, ahem, let you take me out . . . '

He felt furious, then. "I remember you," he said savagely. "For instance the fact that your breasts are sprayed blue; that stuck in my mind. And I remember Mr. McClane's promise that if I remembered my visit to Rekal, Incorporated I'd receive my money back in full. Where is Mr. McClane?"

After a delay—probably as long as they could manage—he found himself once more seated facing the imposing walnut desk, exactly as he had been an hour or so earlier in the day.

"Some technique you have," Quail said sardonically. His disappointment—and resentment—were enormous, by now. "My so-called 'memory' of a trip to Mars as an undercover agent for Interplan is hazy and vague and shot full of contradictions. And I clearly remember my dealings here with you people. I ought to take this to the Better Business Bureau." He was burning angry, at this point; his sense of being cheated had overwhelmed him, had destroyed his customary aversion to participating in a public squabble.

Looking morose, as well as cautious, McClane said, "We capitulate, Quail. We'll refund the balance of your fee. I fully concede the fact that we did absolutely nothing for you." His tone was resigned.

Quail said accusingly, "You didn't even provide me with the various artifacts that you claimed would 'prove' to me I had been on Mars. All that song-and-dance you went into—it hasn't materialized into a damn thing. Not even a ticket stub. Nor postcards. Nor passport. Nor proof of immunization shots. Nor—"

"Listen, Quail," McClane said. "Suppose I told you—" He broke off. "Let it go." He pressed a button on his intercom. "Shirley, will you disburse five hundred and seventy more 'creds in the form of a cashier's check made out to Douglas Quail? Thank you." He released the button, then glared at Quail.

Presently the check appeared; the receptionist placed it before McClane and once more vanished out of sight, leaving the two men alone, still facing each other across the surface of the massive walnut desk.

"Let me give you a word of advice," McClane said as he signed the check and passed it over, "Don't discuss your, ahem, recent trip to Mars with anyone."

"What trip?"

"Well, that's the thing." Doggedly, McClane said, "The trip you partially remember. Act as if you don't remember; pretend it never took place. Don't ask me why; just take my advice: it'll be better for all of us." He had begun to perspire. Freely. "Now, Mr. Quail, I have other business, other clients to see." He rose, showed Quail to the door.

Quail said, as he opened the door, "A firm that turns out such bad work shouldn't have any clients at all." He shut the door behind him.

On the way home in the cab Quail pondered the wording of his letter of complaint to the Better Business Bureau, Terra Division. As soon as he could get to his typewriter he'd get started; it was clearly his duty to warn other people away from Rekal, Incorporated.

When he got back to his conapt he seated himself before his Hermes Rocket portable, opened the drawers and rummaged for carbon paper—and noticed a small, familiar box. A box which he had carefully filled on Mars with Martian fauna and later smuggled through customs.

Opening the box he saw, to his disbelief, six dead maw-worms and several varieties of the unicellular life on which the Martian worms fed. The protozoa were dried-up, dusty, but he recognized them; it had taken him an entire day picking among the vast dark alien boulders to find them. A wonderful, illuminated journey of discovery.

But I didn't go to Mars, he realized.

Yet on the other hand—

Kirsten appeared at the doorway to the room, an armload of pale brown groceries gripped. "Why are you home in the middle of the day?" Her voice, in an eternity of sameness, was accusing.

"*Did I go to Mars*?" he asked her. "You would know."

"No, of course you didn't go to Mars; *you* would know that, I would think. Aren't you always bleating about going?"

He said, "By God, I think I went." After a pause he added, "And simultaneously I think I didn't go."

"Make up your mind."

"How can I?" He gestured. "I have both memory-tracks grafted inside my head; one is real and one isn't but I can't tell which is which. Why can't I rely on you? They haven't tinkered with you." She could do this much for him at least—even if she never did anything else.

Kirsten said in a level, controlled voice, "Doug, if you don't pull yourself together, we're through. I'm going to leave you."

"I'm in trouble." His voice came out husky and coarse. And shaking. "Probably I'm heading into a psychotic episode; I hope not, but—maybe that's it. It would explain everything, anyhow."

Setting down the bag of groceries, Kirsten stalked to the closet. "I was not kidding," she said to him quietly. She brought out a coat, got it on, walked back to the door of the conapt. "I'll phone you one of these days soon," she said tonelessly. "This is goodbye, Doug. I hope you pull out of this eventually; I really pray you do. For your sake."

"Wait," he said desperately. "Just tell me and make it absolute; I did go or I didn't—tell me which one." *But they may have altered your memory-track also*, he realized.

The door closed. His wife had left. Finally!

A voice behind him said, "Well, that's that. Now put up your hands, Quail. And also please turn around and face this way."

He turned, instinctively, without raising his hands.

The man who faced him wore the plum uniform of the Interplan Police Agency, and his gun appeared to be UN issue. And, for some odd reason, he seemed familiar to Quail; familiar in a blurred, distorted fashion which he could not pin down. So, jerkily, he raised his hands.

"You remember," the policeman said, "your trip to Mars. We know all your actions today and all your thoughts—in particular your very important thoughts on the trip home from Rekal, Incorporated." He explained, "We have a tele-transmitter wired within your skull; it keeps us constantly informed."

A telepathic transmitter; use of a living plasma that had been discovered on Luna. He shuddered with self-aversion. The thing lived inside him, within his own brain, feeding,

listening, feeding. But the Interplan police used them; that had come out even in the homeopapes. So this was probably true, dismal as it was.

"Why me?" Quail said huskily. What had he done—or thought? And what did this have to do with Rekal, Incorporated?

"Fundamentally," the Interplan cop said, "this has nothing to do with Rekal; it's between you and us." He tapped his right ear. "I'm still picking up your mentational processes by way of your cephalic transmitter." In the man's ear Quail saw a small white-plastic plug. "So I have to warn you: anything you think may be held against you." He smiled. "Not that it matters now; you've already thought and spoken yourself into oblivion. What's annoying is the fact that under narkidrine at Rekal, Incorporated you told them, their technicians and the owner, Mr. McClane, about your trip—where you went, for whom, some of what you did. They're very frightened. They wish they had never laid eyes on you." He added reflectively, "They're right."

Quail said, "I never made any trip. It's a false memory-chain improperly planted in me by McClane's technicians." But then he thought of the box, in his desk drawer, containing the Martian life forms. And the trouble and hardship he had had gathering them. The memory seemed real. And the box of life forms; that certainly was real. Unless McClane had planted it. Perhaps this was one of the "proofs" which McClane had talked glibly about.

The memory of my trip to Mars, he thought, *doesn't convince me—but unfortunately it has convinced the Interplan Police Agency. They think I really went to Mars and they think I at least partially realize it.*

"We not only know you went to Mars," the Interplan cop agreed, in answer to his thoughts, "but we know that you now remember enough to be difficult for us. And there's no use expunging your conscious memory of all this, because if we do you'll simply show up at Rekal, Incorporated again and start over. And we can't do anything about McClane and his operation because we have no jurisdiction over anyone except our own people. Anyhow, McClane hasn't committed any crime." He eyed Quail. "Nor, technically, have you. You didn't go to Rekal, Incorporated with the idea of regaining your memory; you went, as we realize, for the usual reason people go there—a love by plain, dull people for adventure." He added, "Unfortunately you're not plain, not dull, and you've already had too much excitement; the last thing in the universe you needed was a course from Rekal, Incorporated. Nothing could have been more lethal for you or for us. And, for that matter, for McClane."

Quail said. "Why is it 'difficult' for you if I remember my trip—my alleged trip—and what I did there?"

"Because," the Interplan harness bull said, "what you did is not in accord with our great white all-protecting father public image. You did, for us, what we never do. As you'll presently remember—thanks to narkidrine. That box of dead worms and algae has been sitting in your desk drawer for six months, ever since you got back. And at no time have you shown the slightest curiosity about it. We didn't even know you had it until you remembered it on your way home from Rekal; then we came here on the double to look for it." He added, unnecessarily, "Without any luck; there wasn't enough time."

A second Interplan cop joined the first one; the two briefly conferred. Meanwhile, Quail thought rapidly. He did remember more, now; the cop had been right about

narkidrine. They—Interplan—probably used it themselves. Probably? He knew darn well they did; he had seen them putting a prisoner on it. Where would *that* be? Somewhere on Terra? More likely Luna, he decided, viewing the image rising from his highly defective—but rapidly less so—memory.

And he remembered something else. Their reason for sending him to Mars; the job he had done.

No wonder they had expunged his memory.

"Oh god," the first of the two Interplan cops said, breaking off his conversation with his companion. Obviously, he had picked up Quail's thoughts. "Well, this is a far worse problem, now; as bad as it can get." He walked toward Quail, again covering him with his gun. "We've got to kill you," he said. "And right away."

Nervously, his fellow officer said, "Why right away? Can't we simply cart him off to Interplan New York and let them—"

"*He* knows why it has to be right away," the first cop said; he too looked nervous, now, but Quail realized that it was for an entirely different reason. His memory had been brought back almost entirely, now. And he fully understood the officer's tension.

"On Mars," Quail said hoarsely, "I killed a man. After getting past fifteen bodyguards. Some armed with sneaky-pete guns, the way you are." He had been trained, by Interplan, over a five year period to be an assassin. A professional killer. He knew ways to take out armed adversaries . . . such as these two officers; and the one with the ear-receiver knew it, too.

If he moved swiftly enough—

The gun fired. But he had already moved to one side, and at the same time he chopped down the gun-carrying officer. In an instant he had possession of the gun and was covering the other, confused, officer.

"Picked my thoughts up," Quail said, panting for breath. "He knew what I was going to do, but I did it anyhow."

Half sitting up, the injured officer grated, "He won't use that gun on you, Sam; I pick that up, too. He knows he's finished, and he knows we know it, too. Come on, Quail." Laboriously, grouting with pain, he got shakily to his feet. He held out his hand. "The gun," he said to Quail. "You can't use it, and if you turn it over to me I'll guarantee not to kill you; you'll be given a hearing, and someone higher up in Interplan will decide, not me. Maybe they can erase your memory once more; I don't know. But you know the thing I was going to kill you for; I couldn't keep you from remembering it. So my reason for wanting to kill you is in a sense past."

Quail, clutching the gun, bolted from the conapt, sprinted for the elevator. *If you follow me*, he thought, *I'll kill you. So don't*. He jabbed at the elevator button and, a moment later, the doors slid back.

The police hadn't followed him. Obviously they had picked up his terse, tense thoughts and had decided not to take the chance.

With him inside the elevator descended. He had gotten away—for a time. But what next? Where could he go?

The elevator reached the ground floor; a moment later Quail had joined the mob of peds hurrying along the runnels. His head ached and he felt sick. But at least he had

evaded death; they had come very close to shooting him on the spot, back in his own conapt.

And they probably will again, he decided. *When they find me. And with this transmitter inside me, that won't take too long*.

Ironically, he had gotten exactly what he had asked Rekal, Incorporated for. Adventure, peril, Interplan police at work, a secret and dangerous trip to Mars in which his life was at stake—everything he had wanted as a false memory.

The advantages of it being a memory—and nothing more—could now be appreciated.

On a park beach, alone, he sat dully watching a flock of perts: a semi-bird imported from Mars' two moons, capable of soaring flight, even against Earth's huge gravity.

Maybe I can find my way back to Mars, he pondered. But then what? It would be worse on Mars; the political organization whose leader he had assassinated would spot him the moment he stepped from the ship; he would have Interplan and *them* after him, there.

Can you hear me thinking? he wondered. Easy avenue to paranoia; sitting here alone he felt them tuning in on him, monitoring, recording, discussing . . . He shivered, rose to his feet, walked aimlessly, his hands deep in his pockets. *No matter where I go*, he realized. *You'll always be with me. As long as I have this device inside my head*.

I'll make a deal with you, he thought to himself—and to them. *Can't you imprint a false-memory template on me again, as you did before, that I lived an average, routine life, never went to Mars? Never saw an Interplan uniform up close and never handled a gun?*

A voice inside his brain answered, "As has been carefully explained to you: that would not be enough."

Astonished, he halted.

"We formerly communicated with you in this manner," the voice continued. "When you were operating in the field, on Mars. It's been months since we've done it; we assumed, in fact, that we'd never have to do so again. Where are you?"

"Walking," Quail said, "to my death." *By your officers' guns*, he added as an afterthought. "How can you be sure it wouldn't be enough?" he demanded. "Don't the Rekal techniques work?"

"As we said. If you're given a set of standard, average memories you get—restless. You'd inevitably seek out Rekal or one of its competitors again. We can't go through this a second time."

"Suppose," Quail said, "once my authentic memories have been cancelled, something more vital than standard memories are implanted. Something which would act to satisfy my craving," he said. "That's been proved; that's probably why you initially hired me. But you ought to be able to come up with something else—something equal. I was the richest man on Terra but I finally gave all my money to educational foundations. Or I was a famous deep-space explorer. Anything of that sort; wouldn't one of those do?"

Silence.

"Try it," he said desperately. "Get some of your top-notch military psychiatrists; explore my mind. Find out what my most expansive daydream is." He tried to think. "Women," he said. "Thousands of them, like Don Juan had. An interplanetary playboy—

a mistress in every city on Earth, Luna and Mars. Only I gave that up, out of exhaustion. Please," he begged. "Try it."

"You'd voluntarily surrender, then?" the voice inside his head asked. "If we agreed to arrange such a solution? *If* it's possible?"

After an interval of hesitation he said, "Yes." *I'll take the risk*, he said to himself, *that you don't simply kill me*.

"You make the first move," the voice said presently. "Turn yourself over to us. And we'll investigate that line of possibility. If we can't do it, however, if your authentic memories begin to crop up again as they've done at this time, then—" There was silence and then the voice finished, "We'll have to destroy you. As you must understand. Well, Quail, you still want to try?"

"Yes," he said. Because the alternative was death now—and for certain. At least this way he had a chance, slim as it was.

"You present yourself at our main barracks in New York," the voice of the Interplan cop resumed. "At 580 Fifth Avenue, floor twelve. Once you've surrendered yourself we'll have our psychiatrists begin on you; we'll have personality-profile tests made. We'll attempt to determine your absolute, ultimate fantasy wish—and then we'll bring you back to Rekal, Incorporated, here; get them in on it, fulfilling that wish in vicarious surrogate retrospection. And—good luck. We do owe you something; you acted as a capable instrument for us." The voice lacked malice; if anything, they—the organization—felt sympathy toward him.

"Thanks," Quail said. And began searching for a robot cab.

"Mr. Quail," the stern-faced, elderly Interplan psychiatrist said, "you possess a most interesting wish-fulfillment dream fantasy. Probably nothing such as you consciously entertain or suppose. This is commonly the way; I hope it won't upset you too much to hear about it."

The senior ranking Interplan officer present said briskly, "He better not be too much upset to hear about it, not if he expects not to get shot."

"Unlike the fantasy of wanting to be an Interplan undercover agent." the psychiatrist continued, "which, being relatively speaking a product of maturity, had a certain plausibility to it, this production is a grotesque dream of your childhood; it is no wonder you fail to recall it. Your fantasy is this: you are nine years old, walking alone down a rustic lane. An unfamiliar variety of space vessel from another star system lands directly in front of you. No one on Earth but you, Mr. Quail, sees it. The creatures within are very small and helpless, somewhat on the order of field mice, although they are attempting to invade Earth; tens of thousands of other such ships will soon be on their way, when this advance party gives the go-ahead signal."

"And I suppose I stop them," Quail said, experiencing a mixture of amusement and disgust. "Single-handed I wipe them out. Probably by stepping on them with my foot."

"No," the psychiatrist said patiently. "You halt the invasion, but not by destroying them. Instead, you show them kindness and mercy, even though by telepathy—their mode of communication—you know why they have come. They have never seen such humane traits exhibited by any sentient organism, and to show their appreciation they make a covenant with you."

Quail said, "They won't invade Earth as long as I'm alive."

"Exactly." To the Interplan officer the psychiatrist said, "You can see it does fit his personality, despite his feigned scorn."

"So by merely existing," Quail said, feeling a growing pleasure, "by simply being alive, I keep Earth safe from alien rule. I'm in effect, then, the most important person on Terra. Without lifting a finger."

"Yes indeed, sir," the psychiatrist said. "And this is bedrock in your psyche; this is a life-long childhood fantasy. Which, without depth and drug therapy, you never would have recalled. But it has always existed in you; it went underneath, but never ceased."

To McClane, who sat intently listening, the senior police official said, "Can you implant an extra-factual memory pattern that extreme in him?"

"We get handed every possible type of wish-fantasy there is," McClane said. "Frankly, I've heard a lot worse than this. Certainly we can handle it. Twenty-four hours from now he won't just *wish* he'd saved Earth; he'll devoutly believe it really happened."

The senior police official said, "You can start the job, then. In preparation we've already once again erased the memory in him of his trip to Mars."

Quail said, "What trip to Mars?"

No one answered him, so, reluctantly, he shelved the question. And anyhow a police vehicle had now put in its appearance; he, McClane, and the senior police officer crowded into it, and presently they were on their way to Chicago and Rekal, Incorporated.

"You had better make no errors this time," the police officer said to heavy-set, nervous-looking McClane.

"I can't see what could go wrong," McClane mumbled, perspiring. "This has nothing to do with Mars or Interplan. Single-handedly stopping an invasion of Earth from another star-system." He shook his head at that. "Wow, what a kid dreams up. And by pious virtue, too; not by force. It's sort of quaint." He dabbed at his forehead with a large linen pocket handkerchief.

Nobody said anything.

"In fact," McClane said, "it's touching."

"But arrogant," the police official said starkly. "Inasmuch as when he dies the invasion will resume. No wonder he doesn't recall it; it's the most grandiose fantasy I ever ran across." He eyed Quail with disapproval. "And to think we put this man on our payroll."

When they reached Rekal, Incorporated the receptionist, Shirley, met them breathlessly in the outer office. "Welcome back, Mr. Quail," she fluttered, her melon-shaped breasts—today painted an incandescent orange—bobbing with agitation. "I'm sorry everything worked out so badly before; I'm sure this time it'll go better."

Still repeatedly dabbing at his shiny forehead with his neatly-folded Irish linen handkerchief, McClane said, "It better." Moving with rapidity he rounded up Lowe and Keeler, escorted them and Douglas Quail to the work area, and then, with Shirley and the senior police officer, returned to his familiar office. To wait.

"Do we have a packet made up for this, Mr. McClane?" Shirley asked, bumping against him in her agitation, then coloring modestly.

"I think we do." He tried to recall; then gave up and consulted the formal chart. "A combination," he decided aloud, "of packets Eighty-one, Twenty, and Six." From the vault section of the chamber behind his desk he fished out the appropriate packets, carried them to his desk for inspection. "From Eighty-one," he explained, "a magic healing rod given him—the client in question, this time Mr. Quail—by the race of beings from another system. A token of their gratitude."

"Does it work?" the police officer asked curiously.

"It did once," McClane explained. "But he, ahem, you see, used it up years ago, healing right and left. Now it's only a memento. But he remembers it working spectacularly." He chuckled, then opened packet Twenty. "Document from the UN Secretary General thanking him for saving Earth; this isn't precisely appropriate, because part of Quail's fantasy is that no one knows of the invasion except himself, but for the sake of verisimilitude we'll throw it in." He inspected packet Six, then. What came from this? He couldn't recall; frowning, he dug into the plastic bag as Shirley and the Interplan police officer watched intently.

"Writing," Shirley said. "In a funny language."

"This tells who they were," McClane said, "and where they came from. Including a detailed star map logging their flight here and the system of origin. Of course it's in *their* script, so he can't read it. But he remembers them reading it to him in his own tongue." He placed the three artifacts in the center of the desk. "These should be taken to Quail's conapt," he said to the police officer. "So that when he gets home he'll find them. And it'll confirm his fantasy. SOP—standard operating procedure." He chuckled apprehensively, wondering how matters were going with Lowe and Keeler.

The intercom buzzed. "Mr. McClane, I'm sorry to bother you." It was Lowe's voice; he froze as he recognized it, froze and became mute. "But something's come up. Maybe it would be better if you came in here and supervised. Like before, Quail reacted well to the narkidrine; he's unconscious, relaxed and receptive. But—"

McClane sprinted for the work area.

On a hygienic bed Douglas Quail lay breathing slowly and regularly, eyes half-shut, dimly conscious of those around him.

"We started interrogating him," Lowe said, white-faced. "To find out exactly when to place the fantasy-memory of him single-handedly having saved Earth. And strangely enough—"

"They told me not to tell," Douglas Quail mumbled in a dull drug-saturated voice. "That was the agreement. I wasn't even supposed to remember. But how could I forget an event like that?"

I guess it would be hard, McClane reflected. *But you did—until now*.

"They even gave me a scroll," Quail mumbled, "of gratitude. I have it hidden in my conapt; I'll show it to you."

To the Interplan officer who had followed after him, McClane said, "Well, I offer the suggestion that you better not kill him. If you do they'll return."

"They also gave me a magic invisible destroying rod," Quail mumbled, eyes totally shut, now. "That's how I killed that man on Mars you sent me to take out. It's in my drawer along with the box of Martian maw-worms and dried-up plant life."

Wordlessly, the Interplan officer turned and stalked from the work area.

I might as well put those packets of proof-artifacts away, McClane said to himself resignedly. He walked, step by step, back to his office. *Including the citation from the UN Secretary General. After all—*

The real one probably would not be long in coming.

STUDY AND DISCUSSION QUESTIONS

1 Before you think about the implications of the story about knowledge and skepticism, first ask yourself this: What sort of desire does one have that would prompt going to Rekal, Inc.? Mr. McClane, a Rekal executive and salesman, asks Doug Quail a fascinating question, "So you want to have gone to Mars?" Are subjunctive desires formed in this way coherent and sensible?

2 One way to sharpen that question is by asking another related one: Suppose that you are faced with a choice either to take a $2,500 vacation to Hawaii or to take a, say, $1,250 Rekal "vacation to Hawaii." Imagine that your memories of your time in Hawaii will be identical in both cases. Which do you do? How much more valuable is having a real vacation in Hawaii to you, for example twice as much, five times as much, etc.? Why?

3 Quail seems to desire to escape reality—his job, his wife, and even his planet. In this case Rekal can help. Rekal's guarantee: If you doubt that you went to Mars, Rekal will fully refund your purchase. Descartes also uses a criterion of doubt. How does Descartes employ the concept of doubt in the *Meditations*?

4 This story raises the epistemologically frightening prospect that false memories could be not merely as believable but *more* believable than true memories. When Quail hesitates before purchasing a Rekal treatment, Mr. McClane explains why Rekal memory is better than real memory: "The actual memory, with all its vagueness, omissions and ellipses, not to say distortions—that's second best." How does this create a hurdle for the justification of memory beliefs?

5 After the failed procedure at Rekal, Quail says to his wife, "By God, I think I went. And simultaneously I think I didn't go." By what criteria can Quail determine whether or not he went to Mars? Is introspection useful in justifying his belief? Why/why not?

6 Quail's skepticism is originally restricted to his memory beliefs. But he then realizes that, if his memories are unreliable about having gone to Mars, they are by implication unreliable about his wife. For that matter, his wife's memories may also have been tampered with. This raises a question: Can forms of local skepticism (in this case, skepticism about memory) be contained, or do all or most forms of local skepticism call into question the reliability of other types of beliefs?

2.11 "THE ALLEGORY OF THE CAVE," FROM PLATO'S *REPUBLIC*

Note on Plato's dialogue form: The discussion that follows is a small excerpt about an allegory that Plato uses in *Republic* Book VII to describe his opinions about the nature of human knowledge. In it two characters are having a discussion. These characters are Socrates and Glaucon. Here Socrates is a character in Plato's book, but Socrates was also a real historical person—in fact, Socrates was Plato's philosophy teacher.

Socrates: And now, I said, let me show in a figure how far our nature is enlightened or unenlightened: — Behold! human beings living in a underground den, which has a mouth open towards the light and reaching all along the den; here they have been from their childhood, and have their legs and necks chained so that they cannot move, and can only see before them, being prevented by the chains from turning round their heads. Above and behind them a fire is blazing at a distance, and between the fire and the prisoners there is a raised way; and you will see, if you look, a low wall built along the way, like the screen which marionette players have in front of them, over which they show the puppets.

Glaucon: I see.

Socrates: And do you see, I said, men passing along the wall carrying all sorts of vessels, and statues and figures of animals made of wood and stone and various materials, which appear over the wall? Some of them are talking, others silent.

Glaucon: You have shown me a strange image, and they are strange prisoners.

Socrates: Like ourselves, I replied; and they see only their own shadows, or the shadows of one another, which the fire throws on the opposite wall of the cave?

Glaucon: True, he said; how could they see anything but the shadows if they were never allowed to move their heads?

Socrates: And of the objects which are being carried in like manner they would only see the shadows?

Glaucon: Yes, he said.

Socrates: And if they were able to converse with one another, would they not suppose that they were naming what was actually before them?

Glaucon: Very true.

Socrates: And suppose further that the prison had an echo which came from the other side, would they not be sure to fancy when one of the passers-by spoke that the voice which they heard came from the passing shadow?

Glaucon: No question, he replied.

Socrates: To them, I said, the truth would be literally nothing but the shadows of the images.

Glaucon: That is certain.

Socrates: And now look again, and see what will naturally follow if the prisoners are released and disabused of their error. At first, when any of them is liberated and compelled suddenly to stand up and turn his neck round and walk and look towards

the light, he will suffer sharp pains; the glare will distress him, and he will be unable to see the realities of which in his former state he had seen the shadows; and then conceive some one saying to him, that what he saw before was an illusion, but that now, when he is approaching nearer to being and his eye is turned towards more real existence, he has a clearer vision, — what will be his reply? And you may further imagine that his instructor is pointing to the objects as they pass and requiring him to name them, — will he not be perplexed? Will he not fancy that the shadows which he formerly saw are truer than the objects which are now shown to him?

Glaucon: Far truer.

Socrates: And if he is compelled to look straight at the light, will he not have a pain in his eyes which will make him turn away to take and take in the objects of vision which he can see, and which he will conceive to be in reality clearer than the things which are now being shown to him?

Glaucon: True, he said.

Socrates: And suppose once more, that he is reluctantly dragged up a steep and rugged ascent, and held fast until he's forced into the presence of the sun himself, is he not likely to be pained and irritated? When he approaches the light his eyes will be dazzled, and he will not be able to see anything at all of what are now called realities.

Glaucon: Not all in a moment, he said.

Socrates: He will require to grow accustomed to the sight of the upper world. And first he will see the shadows best, next the reflections of men and other objects in the water, and then the objects themselves; then he will gaze upon the light of the moon and the stars and the spangled heaven; and he will see the sky and the stars by night better than the sun or the light of the sun by day?

Glaucon: Certainly.

Socrates: Last of he will be able to see the sun, and not mere reflections of him in the water, but he will see him in his own proper place, and not in another; and he will contemplate him as he is.

Glaucon: Certainly.

Socrates: He will then proceed to argue that this is he who gives the season and the years, and is the guardian of all that is in the visible world, and in a certain way the cause of all things which he and his fellows have been accustomed to behold?

Glaucon: Clearly, he said, he would first see the sun and then reason about him.

Socrates: And when he remembered his old habitation, and the wisdom of the den and his fellow-prisoners, do you not suppose that he would felicitate himself on the change, and pity them?

Glaucon: Certainly, he would.

Socrates: And if they were in the habit of conferring honours among themselves on those who were quickest to observe the passing shadows and to remark which of them went before, and which followed after, and which were together; and who were therefore best able to draw conclusions as to the future, do you think that he would care for such honours and glories, or envy the possessors of them? Would he not say with Homer,

Better to be the poor servant of a poor master, and to endure anything, rather than think as they do and live after their manner?

Glaucon: Yes, he said, I think that he would rather suffer anything than entertain these false notions and live in this miserable manner.

Socrates: Imagine once more, I said, such an one coming suddenly out of the sun to be replaced in his old situation; would he not be certain to have his eyes full of darkness?

Glaucon: To be sure, he said.

Socrates: And if there were a contest, and he had to compete in measuring the shadows with the prisoners who had never moved out of the den, while his sight was still weak, and before his eyes had become steady (and the time which would be needed to acquire this new habit of sight might be very considerable) would he not be ridiculous? Men would say of him that up he went and down he came without his eyes; and that it was better not even to think of ascending; and if any one tried to loose another and lead him up to the light, let them only catch the offender, and they would put him to death.

Glaucon: No question, he said.

Socrates:

This entire allegory, I said, you may now append, dear Glaucon, to the previous argument; the prison-house is the world of sight, the light of the fire is the sun, and you will not misapprehend me if you interpret the journey upwards to be the ascent of the soul into the intellectual world according to my poor belief, which, at your desire, I have expressed whether rightly or wrongly God knows. But, whether true or false, my opinion is that in the world of knowledge the idea of good appears last of all, and is seen only with an effort; and, when seen, is also inferred to be the universal author of all things beautiful and right, parent of light and of the lord of light in this visible world, and the immediate source of reason and truth in the intellectual; and that this is the power upon which he who would act rationally, either in public or private life must have his eye fixed.

Glaucon: I agree, he said, as far as I am able to understand you.

Socrates: Moreover, I said, you must not wonder that those who attain to this beatific vision are unwilling to descend to human affairs; for their souls are ever hastening into the upper world where they desire to dwell; which desire of theirs is very natural, if our allegory may be trusted.

Glaucon: Yes, very natural.

Socrates: And is there anything surprising in one who passes from divine contemplations to the evil state of man, misbehaving himself in a ridiculous manner; if, while his eyes are blinking and before he has become accustomed to the surrounding darkness, he is compelled to fight in courts of law, or in other places, about the images or the shadows of images of justice, and is endeavouring to meet the conceptions of those who have never yet seen absolute justice?

STUDY AND DISCUSSION QUESTIONS

1 Draw as best you can Plato's cave. As you do so, try to place everything in the allegory in its proper physical place.
2 There are two groups of people in the cave, one of which is tied down and facing a wall. Why do you think Plato portrays them as unable to move, and generally as having many bodily restrictions?
3 What are the prisoners seeing? Specifically, how many levels removed from reality is what they are seeing? Why is Plato using this motif to communicate to his readers about the knowledge of human beings?
4 What is the prisoner's experience upon escaping out of the cave and up to the surface intended to convey? The escapee, now knowledgeable about what lies outside the cave, returns to the cave in order to tell his friends but gets spurned. Is Plato saying that people who achieve true knowledge about reality are often disliked by people who prefer to live lives based upon false beliefs? If so, is he right? Why/why not? Can you identify any examples of this in your culture?
5 Plato's "Allegory of the Cave" can be interpreted in straightforwardly epistemological terms as being about knowledge and reality. But it can also be given a more personal reading. Socially, we often project images of ourselves, consciously or not, which have the effect of distancing others' opinion of us from the way we really are. On the flipside, we often acquiesce in the false or creative self-presentations given us by others. You probably have examples of ways in which you had thought you were "seeing" another person truly, but found out later that he or she was only presenting a "shadow" (or even a "shadow of a shadow") of him or herself to you. What might Plato say about such a form of self-deception?
6 What is the main point Plato is attempting to get across?

2.12 *MEDITATIONS ON FIRST PHILOSOPHY* (I AND II), BY RENÉ DESCARTES

Meditations on first philosophy in which the existence of God and the distinction between the soul and the body are demonstrated

Meditation I: Concerning those things that can be called into doubt

1. Several years have now passed since I first realized how numerous were the false opinions that in my youth I had taken to be true, and thus how doubtful were all those that I had subsequently built upon them. And thus I realized that once in my life I had to

raze everything to the ground and begin again from the original foundations, if wanted to establish anything firm and lasting in the sciences. But the task seemed enormous, and I was waiting until I reached a point in my life that was so timely that no more suitable time for undertaking these plans of action would come to pass. For this reason, I procrastinated for so long that I would henceforth be at fault, were I to waste the time that remains for carrying out the project by brooding over it. Accordingly, I have today suitably freed my mind of all cares, secured for myself a period of leisurely tranquility, and am withdrawing into solitude. At last I will apply myself earnestly and unreservedly to this general demolition of my opinions.

2. Yet to bring this about I will not need to show that all my opinions are false, which is perhaps something I could never accomplish. But reason now persuades me that I should withhold my assent no less carefully from opinions that are not completely certain and indubitable than I would from those that are patently false. For this reason, it will suffice for the rejection of all of these opinions, if I find in each of them some reason for doubt. Nor therefore need I survey each opinion individually, a task that would be endless. Rather, because undermining the foundations will cause whatever has been built upon them to crumble of its own accord, I will attack straightaway those principles which supported everything I once believed.

3. Surely whatever I had admitted until now as most true I received either from the senses or through the senses. However, I have noticed that the senses are sometimes deceptive; and it is a mark of prudence never to place our complete trust in those who have deceived us even once.

But perhaps, even though the senses do sometimes deceive us when it is a question of very small and distant things, still there are many other matters concerning which one simply cannot doubt, even though they are derived from the very same senses: for example, that I am sitting here next to the fire, wearing my winter dressing gown, that I am holding this sheet of paper in my hands, and the like. But on what grounds could one deny that these hands and this entire body are mine? Unless perhaps I were to liken myself to the insane, whose brains are impaired by such an unrelenting vapor of black bile that they steadfastly insist that they are kings when they are utter paupers, or that they are arrayed in purple robes when they are naked, or that they have heads made of clay, or that they are gourds, or that they are made of glass. But such people are mad, and I would appear no less mad, were I to take their behavior as an example for myself.

4. This would all be well and good, were I not a man who is accustomed to sleeping at night, and to experiencing in my dreams the very same things, or now and then even less plausible ones, as these insane people do when they are awake. How often does my evening slumber persuade me of such ordinary things as these: that I am here, clothed in my dressing gown, seated next to the fireplace—when in fact I am lying undressed in bed! But right now my eyes are certainly wide awake when I gaze upon this sheet of paper. This head which I am shaking is not heavy with sleep. I extend this hand consciously and deliberately, and I feel it. Such things would not be so distinct for someone who is asleep. As if I did not recall having been deceived on other occasions even by similar thoughts in my dreams! As I consider these matters more carefully, I see so plainly that there are no

definitive signs by which to distinguish being awake from being asleep. As a result, I am becoming quite dizzy, and this dizziness nearly convinces me that I am asleep.

5. Let us assume then, for the sake of argument, that we are dreaming and that such particulars as these are not true: that we are opening our eyes, moving our head, and extending our hands. Perhaps we do not even have such hands, or any such body at all. Nevertheless, it surely must be admitted that the things seen during slumber are, as it were, like painted images, which could only have been produced in the likeness of true things, and that therefore at least these general things—eyes, head, hands, and the whole body—are not imaginary things, but are true and exist. For indeed when painters themselves wish to represent sirens and satyrs by means of especially bizarre forms, they surely cannot assign to them utterly new natures. Rather, they simply fuse together the members of various animals. Or if perhaps they concoct something so utterly novel that nothing like it has ever been seen before (and thus is something utterly fictitious and false), yet certainly at the very least the colors from which they fashion it ought to be true. And by the same token, although even these general things—eyes, head, hands and the like could be imaginary, still one has to admit that at least certain other things that are even more simple and universal are true. It is from these components, as if from true colors, that all those images of things that are in our thought are fashioned, be they true or false.

6. This class of things appears to include corporeal nature in general, together with its extension; the shape of extended things; their quantity, that is, their size and number; as well as the place where they exist; the time through which they endure, and the like.

7. Thus it is not improper to conclude from this that physics, astronomy, medicine, and all the other disciplines that are dependent upon the consideration of composite things are doubtful, and that, on the other hand, arithmetic, geometry, and other such disciplines, which treat of nothing but the simplest and most general things and which are indifferent as to whether these things do or do not in fact exist, contain something certain and indubitable. For whether I am awake or asleep, two plus three make five, and a square does not have more than four sides. It does not seem possible that such obvious truths should be subject to the suspicion of being false.

8. Be that as it may, there is fixed in my mind a certain opinion of long standing, namely that there exists a God who is able to do anything and by whom I, such as I am, have been created. How do I know that he did not bring it about that there is no Earth at all, no heavens, no extended thing, no shape, no size, no place, and yet bringing it about that all these things appear to me to exist precisely as they do now? Moreover, since I judge that others sometimes make mistakes in matters that they believe they know most perfectly, may I not, in like fashion, be deceived every time I add two and three or count the sides of a square, or perform an even simpler operation, if that can be imagined? But perhaps God has not willed that I be deceived in this way, for he is said to be supremely good. Nonetheless, if it were repugnant to his goodness to have created me such that I be deceived all the time, it would also seem foreign to that same goodness to permit me to be deceived even occasionally. But we cannot make this last assertion.

9. Perhaps there are some who would rather deny so powerful a God than believe that everything else is uncertain. Let us not oppose them; rather, let us grant that everything

said here about God is fictitious. Now they suppose that I came to be what I am either by fate, or by chance, or by a connected chain of events, or by some other way. But because being deceived and being mistaken appear to be a certain imperfection, the less powerful they take the author of my origin to be, the more probable it will be that I am so imperfect that I am always deceived. I have nothing to say in response to these arguments. But eventually I am forced to admit that there is nothing among the things I once believed to be true which it is not permissible to doubt—and not out of frivolity or lack of forethought, but for valid and considered reasons. Thus I must be no less careful to withhold assent henceforth even from these beliefs than I would from those that are patently false, if I wish to find anything certain.

10. But it is not enough simply to have realized these things; I must take steps to keep myself mindful of them. For long-standing opinions keep returning, and, almost against my will, they take advantage of my credulity, as if it were bound over to them by long use and the claims of intimacy. Nor will I ever get out of the habit of assenting to them and believing in them, so long as I take them to be exactly what they are, namely, in some respects doubtful, as has just now been shown, but nevertheless highly probable, so that it is much more consonant with reason to believe them than to deny them. Hence, it seems to me I would do well to deceive myself by turning my will in completely the opposite direction and pretend for a time that these opinions are wholly false and imaginary, until finally, as if with prejudices weighing down each side equally, no bad habit should turn my judgment any further from the correct perception of things. For indeed I know that meanwhile there is no danger or error in following this procedure, and that it is impossible for me to indulge in too much distrust, since I am now concentrating only on knowledge, not on action.

11. Accordingly, I will suppose not a supremely good God, the source of truth, but rather an evil genius, supremely powerful and clever, who has directed his entire effort at deceiving me. I will regard the heavens, the air, the earth, colors, shapes, sounds, and all external things as nothing but the bedeviling hoaxes of my dreams, with which he lays snares for my credulity. I will regard myself as not having hands, or eyes, or flesh, or blood, or any senses, but as nevertheless falsely believing that I possess all these things. I will remain resolute and steadfast in this meditation, and even if it is not within my power to know anything true, it certainly is within my power to take care resolutely to withhold my assent to what is false, lest this deceiver, however powerful, however clever he may be, have any effect on me. But this undertaking is arduous, and a certain laziness brings me back to my customary way of living. I am not unlike a prisoner who enjoyed an imaginary freedom during his sleep, but, when he later begins to suspect that he is dreaming, fears being awakened and nonchalantly conspires with these pleasant illusions. In just the same way, I fall back of my own accord into my old opinions, and dread being awakened, lest the toilsome wakefulness which follows upon a peaceful rest must be spent thenceforward not in the light but among the inextricable shadows of the difficulties now brought forward.

Meditation II: concerning the nature of the human mind: that it is better known than the body

12. Yesterday's meditation has thrown me into such doubts that I can no longer ignore them, yet I fail to see how they are to be resolved. It is as if I had suddenly fallen into a deep whirlpool; I am so tossed about that I can neither touch bottom with my foot, nor swim up to the top. Nevertheless I will work my way up and will once again attempt the same path I entered upon yesterday. I will accomplish this by putting aside everything that admits of the least doubt, as if I had discovered it to be completely false. I will stay on this course until I know something certain, or, if nothing else, until I at least know for certain that nothing is certain. Archimedes sought but one firm and immovable point in order to move the entire Earth from one place to another. Just so, great things are also to be hoped for in succeeding in finding just one thing, however slight, that is certain and unshaken.

13. Therefore I suppose that everything I see is false. I believe that none of what my deceitful memory represents ever existed. I have no senses whatever. Body, shape, extension, movement, and place are all chimeras. What then will be true? Perhaps just the single fact that nothing is certain.

14. But how do I know there is not something else, over and above all those things that I have just reviewed, concerning which there is not even the slightest occasion for doubt? Is there not some God, or by whatever name I might call him, who instills these very thoughts in me? But why would I think that, since I myself could perhaps be the author of these thoughts? Am I not then at least something? But I have already denied that I have any senses and any body. Still I hesitate; for what follows from this? Am I so tied to a body and to the senses that I cannot exist without them? But I have persuaded myself that there is absolutely nothing in the world: no sky, no Earth, no minds, no bodies. Is it then the case that I too do not exist? But doubtless I did exist, if I persuaded myself of something. But there is some deceiver or other who is supremely powerful and supremely sly and who is always deliberately deceiving me. Then too there is no doubt that I exist, if he is deceiving me. And let him do his best at deception, he will never bring it about that I am nothing so long as I shall think that I am something. Thus, after everything has been most carefully weighed, it must finally be established that this pronouncement "I am, I exist' is necessarily true every time I utter it or conceive it in my mind.

15. But I do not yet understand sufficiently what I am—I, who now necessarily exist. And so from this point on, I must be careful lest I unwittingly mistake something else for myself, and thus err in that very item of knowledge that I claim to be the most certain and evident of all. Thus, I will meditate once more on what I once believed myself to be, prior to embarking upon these thoughts. For this reason, then, I will set aside whatever can be weakened even to the slightest degree by the arguments brought forward, so that eventually all that remains is precisely nothing but what is certain and unshaken.

16. What then did I used to think I was? A man, of course. But what is a man? Might I not say a "rational animal"? No, because then I would have to inquire what "animal" and "rational" mean. And thus from one question I would slide into many more difficult ones. Nor do I now have enough free time that I want to waste it on subtleties of this sort. Instead, permit me to focus here on what came spontaneously and naturally into my

thinking whenever I pondered what I was. Now it occurred to me first that I had a face, hands, arms, and this entire mechanism of bodily members: the very same as are discerned in a corpse, and which I referred to by the name "body." It next occurred to me that I took in food, that I walked about, and that I sensed and thought various things; these actions I used to attribute to the soul. But as to what this soul might be, I either did not think about it or else I imagined it a rarified I—know not what—like a wind, or a fire, or ether, which had been infused into my coarser parts. But as to the body I was not in any doubt. On the contrary, I was under the impression that I knew its nature distinctly. Were I perhaps tempted to describe this nature such as I conceived it in my mind, I would have described it thus: by "body," I understand all that is capable of being bounded by some shape, of being enclosed in a place, and of filling up a space in such a way as to exclude any other body from it; of being perceived by touch, sight, hearing, taste, or smell; of being moved in several ways, not, of course, by itself, but by whatever else impinges upon it. For it was my view that the power of self-motion, and likewise of sensing or of thinking, in no way belonged to the nature of the body. Indeed I used rather to marvel that such faculties were to be found in certain bodies.

STUDY AND DISCUSSION QUESTIONS

1 Descartes says that he wants to rid himself of all his former opinions. What about Descartes' historical and cultural context would make him say something like that? Satisfactorily answering that question may require some Internet research on the religious and scientific context in mid-seventeenth century Europe.
2 As a method, Descartes' *method of doubt* is extremely rigorous in its demand to dismiss previously held beliefs. What do you think: is that an unreasonable way to begin a philosophy book, or is it just how such a book should read? Underlying your answer to this question will probably be concerns about whether philosophical inquiry should be objective and unbiased, and about whether total objectivity is possible. Is it?
3 Descartes offers the following analogy for his *method of doubt* on display in his First Meditation. Suppose I want to find the good apples in a barrel full of apples, many of which are bad. Amongst the methods open to me, the most efficient means to my goal is to dump the barrel over and, in one glance, identify the brown, rotten apples from the bright red apples. What does this tell you about Descartes' use of skepticism here—is he a skeptic at heart, or is it merely a means to an end? Why?
4 En route to calling all his beliefs into question, he offers a brief argument in paragraph 3 that his senses are not to be trusted. What is this argument? Is it valid and sound?
5 Paragraph 4 places a constraint or caveat on his enquiries. What is this constraint? What is its philosophical importance?

6 Descartes next mounts an argument from dreams for skepticism. Is Descartes concluding that it is possible that I might always be dreaming, or that it is possible that I might be dreaming right now? What is the philosophical difference between these two conclusions?
7 What beliefs does the dream argument call into question, and what beliefs does it fail to call into question?
8 In paragraphs 10 and 11, Descartes discusses God's role in quelling Descartes' skeptical doubts. Many Christian philosophers throughout the ages also appeal to God as being a guarantor of our knowledge. But in paragraph 12 Descartes subverts the traditional appeal to a good God by replacing it with a deceiving God. From this follows an unholy form of total skepticism. First, how do you think people in Descartes' day would react to this form of skepticism? Second, what beliefs does this "evil demon" skepticism show are unknown that dream skepticism did not show are unknown?
9 In Meditation II, Descartes attempts to answer the skeptical worries from his previous reflections. What is his central response to total skepticism? Why is this response considered a great work of philosophical genius?
10 Think critically about Descartes' alleged solution to his skeptical doubts. Is it as great as people often think it is? What sorts of philosophical problems does his solution have?

2.13 "THE SIMULATION ARGUMENT AND SIMULATION HYPOTHESES," BY ALASDAIR RICHMOND

1. Demons, vats, Sims

Nick Bostrom argues that some plausible ideas about computing and consciousness might imply that reality has a structure rather unlike what we ordinarily imagine. His "Simulation Argument" may sound like Descartes' "Evil Demon" or Putnam's "Brain-in-a-vat" scenarios but really it's quite unlike either, raising interesting metaphysical and epistemic issues all its own. Before tackling Bostrom's argument, let's revisit skepticism and consider some philosophy of mind.

In Descartes' thought experiment, an "Evil Demon" has been feeding all your senses a systematically false model world throughout your life. If no empirical evidence can distinguish "real world" from "evil demon" experiences and all your senses could be simultaneously demon-compromised, maybe all your beliefs about reality are false. If totally demon-managed lives are indistinguishable from "real world" lives, how do you know you aren't an evil demon dupe right now?

Descartes thought some certainties survive radical skepticism. Nothing can deceive him into thinking he exists if he doesn't: *Cogito, ergo sum*. (Descartes also concluded he *was* his mind but *inhabited* his body—he can doubt his physical self exists but not his thinking self. Furthermore, souls are indivisible and so naturally immortal.) Next, Descartes argued, his idea of God proves such a being must exist—neither experience nor finite minds could yield the idea of an omnipotent and perfectly good being. So an omnipotent, omnibenevolent God exists and created (at least some of) our ideas. No such God would create us and our senses, only to leave us in thrall to an evil demon. Hence Evil Demon skepticism must be false. (Many philosophers suspect Descartes inadvertently helped skepticism more than hindered it.)

In Putnam's "Brain-in-a-vat" ("BiV") story, your brain has always lived in a vat and been artificially stimulated with appearances of an external world. Putnam says: thinking you're a BiV is incoherent, because meaning involves preserving correct causal links between signs and things signified. If it's the causal chains that determine their meanings, BiV-words don't mean the same things as our words. A BiV who thinks "I'm a brain in a vat" is *wrong*. How can a BiV falsely assert its own BiV-ness? For a BiV, the word "brain" refers, not to grey matter in a skull, but to a vat-*image* of a brain. (Likewise, for the word "vat" and perhaps for "in.") Words uttered by BiVs are not, and never were, causally linked to brains and vats. A BiV who thinks "I'm a brain in a vat" doesn't mean what we mean by this sentence. Instead, it means (e.g.) "I'm a brain* in a vat*," where "brain*" means "BiV-image of a brain" and "vat*" means "BiV-image of a vat." Thus, a BiV can't truly think "I'm a brain in a vat" because it's a BiV, *not* a brain-image in a vat-image. So, a non-BiV thinking "I'm a brain in a vat" is wrong but a BiV thinking "I'm a brain in a vat" is wrong too. Putnam concludes: a BiV can't *correctly* think "I'm a brain in a vat." (Note newly-envatted brains might retain sufficient causal traces of external objects to let them think "I'm a brain in a vat" with the same reference we possess.) Descartes thought Evil Demon victims massively duped but Putnam thinks BiVs needn't be deceived. A BiV could still gather all available empirical evidence. So, Descartes thought his hypothesis coherent but skeptical; Putnam thinks his hypothesis self-undermining but non-skeptical. However, Bostrom argues for outcomes which should be neither incoherent nor skeptical.

Now some philosophy of mind: *functionalists* think minds stand to brains as computer software does to hardware. Thus, mind is a product of brain-function and consciousness just *is* performing the right sort of operations. Just as software can move between hard-drives without impairment, so human consciousness could function on different substrates. Thus, mind might be "substrate independent," i.e. not dependent on the particular hardware it runs on. Your mind presumably now runs on synaptic hardware but you could think exactly the same thoughts while your mind ran on a properly-configured computer. The contents of your experience don't infallibly reveal the substrate you run on. Your mind could survive changing substrate, and even achieve immortality by adopting an infinite succession of physical substrates. Your mind (*qua* software) could survive forever, even if all its substrates were mortal.

After Weatherson (2003), let conscious simulations be "Sims." Making Sims is beyond us now but our computing capability is far from its theoretical limits. Advanced

("posthuman") beings might already run Sims on vastly better computers than ours. Indeed, there may be more posthuman-run Sims than there are non-Sims.

The Simulation Argument says: functionalists who (a) accept substrate-independence and (b) believe posthumans will run many Sims, should also believe (c) they're probably Sims. Why should (a) and (b) jointly imply (c)? Philosophers often model reasons for belief as guidelines for rational betting. Imagine you believe you have a sign on your back which is either green or red. You can't inspect your own sign but you believe 90 percent of people wear green signs and 10 percent, red. If you've no other information, you should think it 90 percent probable your sign is green. Why? Your evidence concerns relative preponderance of two states. You don't know which is yours but you know you're in one state, and one state is nine times more common than the other. You ought to proportion your degree of belief to what you consider the most populous state. Rational gamblers seek maximal winnings, and "Bet green" promises nine times the winners of "Bet red."

Replace "wearing a green sign" and "wearing a red sign" with "being a Sim" and "being a non-Sim" respectively. Your evidence doesn't determine whether or not you're a Sim. (You haven't received messages from your simulators or glimpsed any programming glitches.) If you think 90 percent of all minds are Sims, your indicated strategy is: give your Sim-hood 90 percent personal probability. Putnam and Descartes only require their hypotheses to be empirically indistinguishable from an external world, and *don't* need any positive evidence for their hypotheses. Instead, Bostrom thinks some assumptions about computing and mind make Sim-hood not only possible but *probable*.

Bostrom says functionalists who accept his argument face a trilemma: (1) posthumans are rare, (2) posthumans run few Sims, or (3) we're probably Sims. Bostrom defends only this trilemma, and does *not* claim that we're probably Sims. The Simulation *Argument* differs from Simulation *Hypotheses*. The former is a probabilistic disjunction; the latter are metaphysical theories about what lies beneath sensory appearances (e.g. we live in a computer simulation). Simulation Hypotheses say our world resembles The Matrix: a virtual realm simultaneously shared by many humans but with an intelligently-controlled computational substratum which we (ordinarily) can't access. (Simulation Hypotheses imply reality is computational at some level, but the reverse implication needn't hold. Reality might have a computational substratum without this substratum being under conscious control. For ease, we'll assume all computationally-generated worlds are simulations.)

Neither Simulation Argument nor Simulation Hypothesis implies the other. You can accept Bostrom's argument but reject your Sim-hood. (You might believe **functionalism** is false or few posthumans exist.) You might reject Bostrom's argument but believe you're a Sim. (Maybe your simulators place a pop-up menu of instructions in your visual field.) The Simulation Argument commits you to the Simulation Hypothesis only if you think Sims out-number non-Sims.

2. Simulation, Doomsday and anthropic reasoning

Compare Bostrom's argument with Brandon Carter and John Leslie's "Doomsday Argument" (see Leslie 1998). Both use the "Anthropic Principle," coined by Carter to express how our nature as observers relates to the physical conditions we observe. As only certain kinds of physical conditions support observers, only certain kinds of conditions get observed. (I bet you're reading this in a temperate oxygen-bearing environment and not inside a star.) Doomsday applies anthropic reasoning to our position in human history, not our spatial or physical location.

Doomsday suggests we should give low probability to human population expanding (or holding steady), and high probability to our population irreversibly declining. Perhaps 10 percent of all people who have ever lived are alive now, so our *c.*6 billion contemporaries have birth-ranks *c.*60 billion. If this 10 percent is a significant fraction of everyone who ever *will* live (i.e. humanity nears extinction), our current position is fairly unexceptional. (Almost all birth-ranks will equal, or be less than, ours.) If Doom is deferred and most people live after us, we're unusually early humans. (Birth-ranks below 60 billion are then atypical, because most birth-ranks will exceed 60 billion.) If we should prefer explanations that maximize our location's probability, we should assume imminent extinction is likely. (Strictly, "Doomsday" is misnamed, as the argument is compatible not only with our extinction but our evolving into something else. Both Doomsday and Simulation arguments have disjunctive conclusions. For a fictional treatment of Doomsday, see Baxter (2000).)

Leslie illustrates Doomsday thus: (i) Your name is written on a slip of paper and placed in an urn. Assume the urn either holds ten names or a million names. Names are drawn randomly from the urn, and aren't replaced. If your name is drawn third, this datum favors the "ten name" hypothesis. (ii) You draw balls randomly from an urn (without replacement) and the first 100 balls you draw are all green. You have two hypotheses: H1—the urn holds 1,000 green balls; H2—the urn holds 100 green balls and 900 red ones. If you assume a random draw, you should favor H1, because H1 confers higher probability on your data. If H2 is true, your sample is atypical, containing the only green balls from an urn 90 percent full of red balls.

Like all anthropic arguments, Doomsday assumes we ought to favor whichever hypothesis makes our location as observers probable. Bostrom accepts this assumption but rejects Doomsday. He says knowing our approximate birth-ranks (i.e. that we live *c.*2007 CE) means we can't treat ourselves as random humans. We don't possess any direct evidence as to whether or not we're Sims but we do have abundant population data.

3. Simulation hierarchies

Bostrom thinks our creating Sims would suggest (1) posthumans aren't rare and/or (2) posthumans run many Sims. If/when we create Sims, this would tell heavily against two options in Bostrom's trilemma and suggest we're probably Sims. Presumably properly

simulating a civilization would involve simulating all its Sims. If functionalism is correct, simulated Sims would themselves be conscious. So, if we start running Sims, we should think we probably live in a nested hierarchy of simulating simulators (like Sawyer's auto-homicide hierarchy in "Iterations').

Each level in a Sim-hierarchy can effectively be its own world with its own laws. Presumably, simulators at the bottom bear the costs of simulating all higher layers. If basement simulators have only finite resources, presumably there can be only finite Sim-hierarchies. Even if only finitely many Sim-levels can exist, a hierarchy might hold many levels and each level, many Sims. (If the basement simulator is a god-like being of infinite resources then reality could be infinitely-layered and extend indefinitely.)

Causal and evidential asymmetries will exist between lower and higher levels. Higher levels are causal descendents of lower ones, but not vice-versa. Levels must know about any simulations they run, but needn't know if anyone is simulating them. Bostrom requires these asymmetries, e.g. so we can't dismiss our Sim-hood because we see no-one simulating us. Lower levels might effectively act as gods to those above them, e.g. monitoring and controlling their lives and physical environments. Bostrom (2003a: p. 254) thinks the mere *possibility* of a multi-layered reality might induce moral behavior in all levels. Any level might reasonably fear detection and punishment of its moral failings from levels lower down. If no level can be sure that it is the bottom, all levels have an incentive to treat other beings well.

4. Skepticism: Truman Burbank, Glaroons and Cthulhu

Can you believe you're a Sim without being metaphysically confused? Could everyday knowledge survive such a belief? It seems as if Sims occupy the same shaky footing as BiVs or Evil Demon dupes, and that Simulation Hypotheses are radically skeptical. Sims, BiVs and Evil Demon dupes all inhabit artificial worlds where appearance and reality can diverge.

However, this skeptical appearance may be deceptive. If you're an Evil Demon dupe, there may be no reliable memories in your mind, no other people, no physical objects as you imagine them, no space or time, no mathematical truths—all could be Demon-spawned illusions. Maybe you and the Demon exhaust the sum total of existence. However, if Simulation Hypotheses are correct, there can still be other human minds in a shared (computational) space. Simulation Hypotheses may just be novel stories about what underlies the objects we perceive. Sims may preserve all the correct causal links to computers, other minds etc., necessary for meaningfully thinking "I'm a Sim." So, Sims need be neither metaphysically incoherent (like BiVs) nor massively deceived (like Evil Demon dupes). Indeed, we might even have good reason to think Sim-hood is our current state.

Fictional scenarios might illustrate which Sim-worlds are, and aren't, radically skeptical. Consider Truman Burbank, titular hero of the film *The Truman Show* and the 24-hour reality TV show within the film. Truman lives in a perpetually stage-managed

world, dedicated solely to keeping him functioning and telegenic. He is the victim of a massive conspiracy and his view of reality is comprehensively skewed. (His supposed friends and neighbors are all actors, charged with keeping him happy and slipping a little product-placement into the show under his unsuspecting nose.) Bostrom thinks that while "Truman worlds" could be governed by exactly the same physical laws we think govern our world, such worlds are nonetheless genuinely skeptical—no one is trustworthy, all is not as it appears. "Truman world" engineering might become necessary if a Sim-world malfunctioned and threw up glitches its Sims could observe. Simulators might intervene to erase memories of glitches or encourage Sims to dismiss glitches as hallucinations. So, even if your local simulation failed, you wouldn't necessarily remember such failures and so you can't infer such things don't happen just because you don't recall them. A lack of observed glitches in our environment doesn't prove we aren't Sims but this suggestion buys the Simulation Hypothesis compatibility with experience at a skeptical price.

The protagonist of Robert Heinlein's story "They" is in a mental institution, having concluded that neither everyday, scientific nor religious explanations for human existence make any sense. He believes the inconsistencies of human behavior prove we're being manipulated by other beings with agendas of their own. In fact, he's right; his world is studded with beings called "Glaroons," who continually scrutinize him disguised as doctors and nurses. So, Glaroons belong to the same space-time manifold as us but control the fundamentals of perception. "Glaroon" worlds too seem deeply skeptical—your perceptions and environment are continually monitored and controlled by hidden powers.

In Heinlein's "The Unpleasant Profession of Jonathan Hoag," Hoag returns from work every evening with no recollection of what his job entails. Hoag gradually discovers he is not human but merely inhabits a human shell. This immersion has occurred because Hoag is a critic, making an in-depth assessment of a work of art, i.e. our world and all its contents. Completing his task obliges Hoag to make adjustments to our reality. So our world is an artifact whose laws bind its inhabitants, but not its creators. Hoag-worlds don't seem *radically* skeptical—in Hoag-worlds, our physical laws are real and (normally) apply everywhere we can observe but such laws can be over-ridden by the world's makers.

In H. P. Lovecraft's "Cthulhu Mythos" stories, our world is an orderly island in an otherwise chaotic universe, beset by alien and/or extra-dimensional beings. The laws governing our matter, geometry and colors are real hereabouts but needn't obtain everywhere. Lovecraft worlds exemplify what D. M. Armstrong (1975: p. 104) called "cosmic epoch" skepticism; a non-radical skepticism whereby our local laws are correct but not spatio-temporally invariant.

Simulation could be Hoagian, Glaroonian or Lovecraftian. In a Matrix, our (local) physics may be real qua our shared perceptual world but it can be over-ridden and needn't apply to all existence. In the *Matrix* trilogy, Neo grows from passive consumer of Sim-reality to Hoagian world-shaper, (like Morpheus and Trinity). Agent Smith's hostile editing of Sim-reality is Glaroonian. The trilogy explains the Matrix's existence in Glaroonian/Lovecraftian fashion: hostile Machines built the Matrix to hold our minds in thrall so they can use our bodies as batteries. (Is thermodynamics different in reality-at-large?) However, there's no *intrinsic* reason why simulators need be hostile. Imagine a benign Matrix, (e.g.)

wherein benevolent powers feed human survivors of nuclear war a simulated world more pleasant than scorched reality

Simulation can have theological dimensions. Consider the problem of evil, i.e. reconciling God's omnipotence and omnibenevolence with the existence of suffering. Bostrom offers this (admittedly far-fetched) Sim-answer to the problem of evil: "There is no suffering in this world and all memories of suffering are illusions," (2003a: p. 254). Besides inciting skepticism about the past and other minds, this answer faces the problem that delusory memories of suffering are still evils. Is it consoling to think suffering is genuinely painful but has a delusory cause? (Offered this Matrix "solution" to suffering, one might want neither red nor blue pills, but a refund.) Generally, philosophical accounts of evil don't dismiss suffering as unreal but argue that suffering is a real but inescapable accompaniment of greater goods, e.g. personal growth, moral autonomy or free will. In *The Matrix: Reloaded*, Neo asks the Matrix's Architect why the Matrix is so bleak if it was built to keep humans entertained. The Architect replies: original Matrices were Utopian but their inhabitants found perfection boring and became ungovernable. Sim-humans seemingly need pain in order to keep their mental wheels turning.

5. Simulationism as revisionary metaphysics

David Chalmers thinks the Matrix hypothesis is not a *skeptical* hypothesis that undermines our beliefs about the physical world but a *metaphysical* hypothesis about what underlies the physical world. Chalmers compares the Matrix Hypothesis to Berkelian immaterialism and quantum theory—neither denies the existence of familiar objects but offers a new account of what reality is. In quantum theory, events can lack causes, a particle's position and velocity cannot simultaneously be known with arbitrary precision (cf. Heisenberg's "uncertainty principle'), and objects may exist in "superpositions" of contradictory states, (cf. Schrödinger's cat). Philosophy too has a long history of revisionary metaphysics. Like his mentor Parmenides, Zeno of Elea (*fl. c.*370 BCE) held that all reality is One, and hence motion and change are impossible. Thomas Hobbes (1588–1679) argued, contrary to Cartesian dualism, that all existence is material, including thoughts, souls and even God. J. M. E. McTaggart (1866–1925) thought time cannot exist and all temporal appearances are false.

According to George Berkeley (1685–1753), only minds and their ideas exist. Non-thinking things exist if (and only if) they are presented to consciousness. Berkeley thought his metaphysics was *not* revisionary, because commonsense would agree with him that *Esse est percipi* ("To be is to be perceived"). Berkeley accepted the existence of chairs, tables and mountains but rejected the idea of Matter, conceived as an imperceptible material substratum beneath what we perceive and which somehow causes our ideas. Thus, a tree is the sum of all the (actual and possible) perceptions of it that exist in minds. Beyond our perceptions, no extra imperceptible "real" tree exists. Does the tree outside disappear if I close my eyes? No, Berkeley says, the tree still appears to God's omnipresent perception and won't disappear even if all finite minds stop perceiving it. Berkeley saw

Nature as not unlike a vast theistic simulation or virtual reality, a gigantic language of signs orchestrated by God for our benefit.

Chalmers says Simulation Hypotheses offer a kind of Cartesian mind/body dualism—our selves might be confined to a Sim-world now but could survive uploading into another world. On this view, our consciousness might genuinely be distinct from any of the (simulated) physical processes we can observe. But unlike Cartesian dualism, simulation can allow a strictly materialistic immortality—Chalmers' selves don't need any *immaterial* substratum to be able to survive Sim-death indefinitely.

Immanuel Kant (1724–1804) distinguished between the "phenomenal realm" and "noumenal realm" (i.e. respectively the world as we perceive it and the world as it is in itself). He thought space and time were real properties of the phenomenal world, but that the noumenal world is not spatio-temporal. Space and time are inescapable categories of our grasp of the world but don't apply to reality-at-large. Chalmers thinks the relation between simulators and Sims resembles that between Kantian noumena and phenomena. Simulators determine the structure of Sim-reality but need not themselves be bound by that structure. Simulation Hypotheses mean our level's noumenal realm may be a phenomenal realm to another level. Our simulators may have noumenal attributes for *us* (e.g. metaphysical foundations beneath our perceptions that can't be accessed directly), but they're presumably phenomena to *themselves*. There may be a *regress* of levels; each level seeming noumenal to its Sims but merely phenomenal to itself and its simulators.

The *Matrix* trilogy gives reality only two levels: the Matrix level and the Zion/Machine level. However, for all its inhabitants can tell, the latter may itself be a simulation run by still deeper levels, which in turn . . . Where the trilogy offers a simple, stable Sim-hierarchy (early on, we learn how many levels there are and which level is which), Simulation Hypotheses allow a regress of levels. David Cronenberg's 1999 film *Existenz* and Christopher Priest's novels *A Dream of Wessex* (1978) and *The Extremes* (1998) brilliantly portray disturbing regressive virtual realities. Many of Philip K. Dick's stories are virtuoso exercises in appearance/reality discords, but see particularly his novels *Time Out of Joint* (1959), *Martian Time-Slip* (1964) and *Ubik* (1969).

6. Skepticism redux?: kinds of Sim, history and truth-tracking

Can skepticism be kept at bay in a Sim-world? It might depend on the kind of simulation we live in and how we come to be in it. Eight arguments (of varying plausibility) follow, some of which suggest that Sim-hypotheses can be skeptical.

(i) If we live in a Matrix, the natural laws we observe may be contingent on our world's programming. Hence none of the laws we think really govern reality may actually do so. How do we know we're generalizing correctly from our experience? Maybe in unsimulated reality-at-large, cold makes water boil, and only in Sim-worlds does cold accompany freezing. However, the problem of how to generalize correctly from experience is the traditional problem of *induction*. Induction is extrapolating from data to a conclusion that goes beyond what the data strictly imply. You might observe a million white swans

and conclude inductively "All swans are white." However, the next swan you observe could prove to be black. Future observations need not conform to past observations and even well-supported inductions can go astray.

David Hume (1711–1776) asked what *rational* warrant we have for induction. Saying the future must resemble the past because it has always done so before is to argue in a circle, assuming the inductive principle at stake. Hume thought induction couldn't be given non-circular justification but was nonetheless inescapable if we want to survive. We can't prove the future will resemble the past or the unobserved will resemble the observed. We have no *logical* guarantee that Nature won't spring nasty surprises or that the laws we think govern the world aren't plain wrong. However, Hume says, we must assume the unobserved will resemble the observed, otherwise we can't live. With respect to induction, Sims and non-Sims are in the same boat. Sims can't prove that reality-at-large must conform to their expectations but neither can non-Sims. Any argument that Sim-hypotheses are skeptical must show that Sims face skeptical problems non-Sims don't. *General* skeptical challenges don't threaten any particular hypothesis.

(ii) Maybe the Sim-hypothesis is skeptical because it makes unfalsifiable claims, i.e. no empirical evidence can prove we aren't Sims. However, many non-skeptical meta-physical claims seem unfalsifiable. I can imagine observations that could falsify the hypothesis "There is a rhinoceros under my desk"; it's harder to imagine how to falsify the hypotheses "Other people exist" or "There is an external world." The Sim-hypothesis may be unfalsifiable but this doesn't necessarily make it skeptical.

(iii) In one sense, Sim-life is like BiV-life: Sim-language will probably have different referents from ours. In the Matrix, Neo has hair but in reality-at-large he's bald—so for Neo, "virtual hair" stands in for our "hair." "Hair" in the Matrix means (e.g.) "flexible stuff that grows out of follicles" but at the next level up, "hair" means (e.g.) "sub-routine X." And so it goes, with each level having a different sub-routine that acts as place-holder for the thing experienced in the Matrix as hair. However, because the inhabitants of the Matrix share a communal and stable virtual world, they can possess a stable, rule-governed language—it just may not have the same referents as ours. Sims needn't be BiVs.

(iv) Our simulators might simulate many alternative histories, so the different levels resemble the branching worlds hypothesized by physicists. (Although quantum worlds are effectively causally isolated from each other once they've branched, whereas simulators are causally linked to the worlds they simulate.) Might a multi-level Sim reality lead to *historical* skepticism? Sims like us might wrongly think they occupy a 21st century pre-Sim period. However, while such uncertainty induces some skeptical unease, it might not support radical skepticism. We may be unsure as to our place in history-at-large but still confident about our place in local (simulated) history. I'm sure I live roughly 90 years after *a* First World War, but my confidence in this proposition needn't falter if not all my contemporaries share this property. How disturbing we find historical divergences between our world and reality-at-large may have a lot to do with what our simulators are like and how long our simulation lasts. Nothing seems more dreadful than this (simulated) life ceasing and our waking up to (e.g.) a Nazi-dominated world. Fortunately, we've no reason to think (a) reality-at-large is totalitarian or (b) our present simulation is transitory. However, even

benevolent simulation might not prevent reality-at-large from striking us as very strange indeed. The Matrix trilogy barely hints at how far reality and appearance might diverge—e.g. in or out of the Matrix, Neo looks like Keanu, space has three dimensions and time runs one way. Philosophy, physics and fiction have all entertained more exotic possibilities.

(v) Consider *holographic* versus *depth* simulation—the former encodes details about objects' surfaces but the latter simulates underlying microstructure too. Holographic simulation suffices for most purposes. Why waste computing resources simulating light-rays passing between chair and eye if you can implant chair-grasping cognitive results directly into Sim-consciousness? If we live in a surface-only holographic simulation then physical objects and their causal powers drop out of the equation. Considerations of economy favor holographic simulation over depth-simulation and thus prompt a measure of skepticism. Thus, most Sims only perceive surface simulations, whose objects have a discontinuous existence. Holographic simulation seems Berkelian: objects exist fully only when perceived and otherwise exist only as potentially perceived.

(vi) Putnam's argument that certain BiV-beliefs are incoherent needn't work for recently-envatted brains. Likewise, short-term immersion in Sim-worlds might raise skeptical worries. Dainton (2002) talks of "day-simulations": short-term rewirings of our neural network that give us a transitory full-immersion visit to another life. Just as Sim-hood and non-Sim-hood are empirically indistinguishable, likewise there may be no detectable differences between being lifetime Sims and daytripper Sims.

How different can parent and simulated selves be before they become different persons? The self typing this likes far-fetched British 1960s TV spy series *The Avengers*. If I chose from a menu of simulated selves, how come I got a "nondescript lecturer" module and not a "John Steed, debonair secret agent" module? If I'm now a daytripper-Sim for my parent self, my parent self's tastes differ from mine. Skepticism looms because I can't safely reason from my Sim-tastes to "my" tastes in reality-at-large. Maybe when this module runs out, my current preferences will too and I'll find myself inhabiting not merely a history but a *self* this self wouldn't recognize. Perhaps Sim-me reads about Lovecraft monsters because (one level down) I am a Lovecraft monster. (Sim-me devoutly hopes not. Monster-me may rejoice in its unspeakable condition but fear that it's pitifully human another level down.) If I'm not a Sim, a BiV or dreaming, only a huge disruption of natural laws could make human-type experiences be succeeded by Cthulhu-type experiences.

If I'm a research-Sim, run to help check a hypothesis about life in 2007, why *this* particular (pleasantly unexciting) existence? Dainton suggests posthumans might be most interested in the period when simulation began to be discussed. But why should this period be more interesting than any other? Perhaps an observer selection-effect operates instead, so wondering if you're a recreational Sim makes it *less* likely that such is your condition. Any life interesting enough to be a popular choice of Sim-module probably wasn't spent pondering the Simulation Argument.

(vii) Consider Gettier cases, i.e. justified true beliefs which fail to be knowledge. Robert Nozick tried defeating Gettier cases with a "truth-tracking" theory of knowledge. Nozick's theory replaces JTB's justification requirement with two conditionals. Thus, a proposition ("P") counts as knowledge if it fulfils four conditions:

1 P is true.
2 You believe P.
3 If P stopped being true, you would stop believing P.
4 If P were to remain true, you would still believe P.

Thus, Nozick says a belief qualifies as knowledge if it is sensitive to the truth of the proposition believed. Here we might get a diagnosis of why the BiV hypothesis seems skeptical. If I was a BiV, I needn't believe I was a BiV, so my beliefs about BiV-hood needn't track the truth. Likewise, if I were a newly-embodied ex-BiV, I might not notice any change and keep thinking I was a BiV. Like beliefs in BiV-hood, beliefs in Sim-hood needn't track truth. If Sim-hood is indistinguishable from non-Sim-hood, you could move between them and not notice. (Substrate-independence makes this a real possibility.) Bostrom requires such failure of truth-sensitivity, because his argument must treat Sim-hood and non-Sim-hood as plausibly indistinguishable. If failing to track truth makes hypotheses skeptical, the Simulation Argument uses a skeptical hypothesis after all.

(viii) While Simulation Hypotheses needn't be nightmarish, they still do violence to ordinary metaphysics. Revisionary metaphysicians can (broadly) make two claims: (i) the revisionary theory has explanatory advantages over its competitors, and/or (ii) competing theories are incoherent. (Quantum theory's unparalleled empirical success reconciled scientists to its metaphysical oddities; Zeno and McTaggart thought motion and time respectively would have to combine logically incompatible properties and therefore *couldn't* exist.) Neither claim seems compelling in the Sim case. By hypothesis, Sim-life is empirically indistinguishable from non-Sim life, and no one has proposed any reason why the non-Sim hypothesis is incoherent.

However, the epistemologist's "total evidence requirement" says we should take into account *all* available evidence before weighing up competing hypotheses. We can't tell whether or not we're Sims but we may have some evidence that bears on our Sim-hood. Bostrom says if we run Sims, we should be more inclined to believe we're Sims ourselves and that reality is a Sim hierarchy. Presumably, such a hierarchy must build up a from a primary "basement" level of non-simulated simulators and terminate in a unique layer of non-simulating Sims. If we are non-simulating Sims, we presumably occupy this unique terminus. However, the more levels reality has, the less likely such a location seems. If we want our level of reality to appear likely, we ought to think that reality doesn't boast other (undetectable) levels below ours. If Bostrom's argument induces a probability-shift towards our Sim-hood, thinking about our non-simulating location can induce an opposing probability-shift.

7. Summary and conclusions

Although not entirely convinced by Bostrom's conclusions, I find his Simulation Argument a fascinating addition to the metaphysical and epistemic literature on computation, probability and even reality. It's also one of the most interesting products to date of the

two-way traffic of ideas between philosophy and science fiction. It can't be repeated often enough that you might find an argument of lasting interest and importance even if you find yourself not in complete sympathy with its conclusions. If nothing else, Bostrom's argument can serve as a fine ice-breaker and intuition-tester for lectures on such diverse topics as probability, induction, knowledge and the anthropic principle.

We can certainly grant Bostrom that his argument differs significantly in its methods and conclusions from Cartesian or BiV hypotheses. However, with that granted, we might still be wary of the claim that functionalism plus a belief in achieved posthumanity should make us think that we're Sims. Updating our beliefs in favor of Sim-hood is no small revision of our beliefs and we would need to put such an updating in a broader epistemic context before we contemplated such a step.

Although not all Simulation Hypotheses are radically skeptical, many have skeptical implications, and compelling reasons for entertaining such revisionary metaphysics aren't yet forthcoming. However, there doesn't seem to be any incoherence in the Simulation Argument, or Simulation Hypotheses. How many levels reality contains, and how they relate to one another, are still open questions. Plenty of hitherto-unobserved but perfectly imaginable *Matrix*-style phenomena may yet be found to occur, and so drive us (or our descendents) towards entertaining something like the Simulation Hypothesis. Maybe in the various realms of quantum mechanics, cosmic fine-tuning and human self-awareness we have already encountered processes that only a computational model of reality can explain.

STUDY AND DISCUSSION QUESTIONS

1. What is the simulation argument's conclusion? Do you find it plausible? What sort of argument is it that Bostrom offers? In other words, is it deductive or inductive or an argument from elimination?
2. What assumptions must the advocate of the argument make about (i) computing, (ii) the nature of consciousness and (iii) how to derive beliefs from probabilities?
3. How does the simulation argument differ from the simulation hypothesis? Does the world portrayed in the films of the *Matrix* trilogy provide a good model of the simulation hypothesis?
4. The simulation argument might have something in common with doomsday and anthropic arguments. What similarities do these arguments show and how do they differ?
5. David Chalmers thinks that the simulation hypothesis isn't a skeptical hypothesis. What does Chalmers think the simulation hypothesis is and do you agree? What reasons might there be for thinking the simulation hypothesis is skeptical after all?

6 Bostrom suggests his argument might have moral and even theological implications. What might these implications be? Do you find them convincing? What other implications (philosophical or otherwise) might Bostrom's argument have?

7 How, if at all, might life as a Sim differ from life as a normally-embodied mind? Would being a Sim be like being a Putnam-style brain-in-a-vat or a victim of Descartes' evil demon? If not, how do these scenarios differ?

8 If you could choose to live as a Sim, would you do so? If not, why not? I invite you not to dismiss this question as having an obvious answer. Think about it: if you are in a Sim world—one that offers you some control over your virtual reality, wouldn't you find more enjoyment living in it? You can eat exactly what you want, have the job you really want (or not have one at all), have the sexual experiences you really want, etc., all at the merest whim.

2.14 BIBLIOGRAPHY AND RECOMMENDED READING

Science fiction

*Baxter, Stephen. *Time: Manifold 1*. London: HarperCollins, 2000. Nominated for the 2000 Arthur C. Clarke award, this highly philosophical novel engages in topics including the doomsday argument and genetic engineering. People attempt to change time, and events in it, in order to prevent the extinction of post-humanity.

Dick, Philip K. *Eye in the Sky*. New York: Vintage Books; reprint ed., 2003. The mental lives of several people are cast into the mind of a single person through an accident at a particle accelerator. Dick plays with issues of solipsism, as well as McCarthyite political surveillance in this heady novel.

——. *Martian Time-Slip* (1964, many subsequent editions). Mental illness, the physics of time and the perception of alternate realities blend thematically as colonists on Mars attempt to become self-sufficient.

*——. *A Scanner Darkly*. New York: Vintage Books; reprint ed., 1977. Drawn from a verse in 1 Corinthians 13, the scanner which is seen through darkly refers to a holographic recording device allowing people to view their own lives. The characteristic blending of the real and unreal occurs on many levels in this novel, beginning with the protagonist, who is a double agent, a hallucinogenic drug addict and a user of scanners. Indeed, the drug, Substance D, in effect separates the two hemispheres of the brain, rending one person into two. The novel is semi-autobiographical.

——. *The Three Stigmata of Palmer Eldritch*. Gollancz, reprint ed., 2003. Colonists must take a concoction of drugs to acclimate to the Martian world and preserve their sanity. Eldritch supplies them with a new drug which not only changes the perception of reality, but also, apparently, recreates reality.

——. *Time Out of Joint* (1959, many subsequent editions). Ragle Gumm slowly realizes that he is living in a virtual reality environment as objects he sees fade in and out of existence. He realizes that there is much more to his world than he had thought, à la *The Truman Show* (rental). But, characteristic of Dick's frenetic plotting, this is but is much more than a story about appearance/reality distinctions. Themes of military exploitation of an innocent civilian for success on the battlefield predate and influence subsequent novels, like *Ender's Game*.

——. *Ubik* (1969, many subsequent editions). The protagonist is engaged by a corporation to do some espionage regarding telepaths and their influence on business. He is double-crossed, and put in an unusual form of cognitive stasis.

——. "We Can Remember It for You Wholesale." *The Magazine of Fantasy and Science Fiction*. April 1966.

Galouye, Daniel. *Dark Universe*. Gollancz, new ed., 1961. Rich and detailed Hugo nominee is about herbivorous human beings who inhabit another world in the conditions quite like Plato's Cave. They live in a lightless environment, which leads them to worship light, and must indirectly infer everything about their surroundings.

Haldeman, Joe W. *A Mind of His Own.* Analog, 1974. Focuses on problems with distinguishing dreams from reality.

*Heinlein, Robert A. "They" and "The Unpleasant Profession of Jonathan Hoag." *The Fantasies of Robert A. Heinlein* New York: Tom Doherty Associates, 1999.

Jaeger, Muriel. *The Man with Six Senses*. London: Hogarth Press, 1927. Especially philosophical novel in which the protagonist attempts to craft a new sense.

Jeter, K. W. *Noir*. Bantam Spectra, 1999. A misanthropic L.A. detective, given the cringe-worthy name of McNihil (literally, "son of nothing"), gets his eyes surgically "enhanced" so that he sees everything as though it were a black and white noir film from the 1930s. The plot picks up on the theme about perception and knowledge as it sinks into a murder mystery involving a runaway suspect who has fled into virtual reality.

Kaul, Fedor. *Contagion to this World*. Bles, 1933. Imagine a world in which everyone got amnesia and lost all memory knowledge. A rogue scientist creates just such a plague, and the result is catastrophic for our species.

Kepler, Johannes. *Somnium: The Dream, or Posthumous Work on Lunar Astronomy*. Trans. Edward Rosen. Dover Publications, 2003. Kepler uses the medium of what we now know as science fiction as the ideal vehicle for his Galilean understanding of the solar system. The story relates a journey to the moon and, from that site, the observation of the solar system and its celestial bodies. Often considered the first work of the genre.

Lovecraft, H. P. *The Call of Cthulhu and Other Weird Stories.* S. T. Joshi, ed. New York: Penguin, 2001.

Priest, Christopher. *A Dream of Wessex*. London: Faber, 1977. A group of British futurists undertake an experiment to create a virtual-reality society. Once in, members forget who they had been by undergoing a form of hypnosis. In addition to working with themes of the perception and reality of space and time, the novel includes intriguing ethical dimensions that lead to thoughts of John Rawls and the "veil of ignorance."

——. *The Extremes*. St. Martin's Press, 1999. 393pp. Virtual reality technology, branded "Extreme Experience," is used extensively by an FBI agent who is attempting to reconstruct and understand two simultaneous massacres, one in Texas, in which her

husband was killed, and the other in England. She travels to England to investigate any connection, and begins using ExEx extensively. Priest draws this to his characteristically stunning conclusion by revealing connections between technology and consciousness.

Sturgeon, Theodore. *Not Without Sorcery*. New York: Ballantine, 1948. "The Ultimate Egoist," contained in this collection, explores ideas in skepticism and solipsism.

Swanwick, Michael. "Scherzo with Tyrannosaur." *Asimov's Science Fiction* 282: July, 1999. A formal dinner, a fascinated boy glued to the window, packs of dinosaurs for entertainment, and much more.

*Zelazny, Roger. *The Dream Master*. New York: Ace, 1966. Based on the Nebula winning "He Who Shapes" (*Amazing*, 1965). *The Dream Master* portrays a psychologist who enters his patient's awareness and "shapes" vivid dreams as part of the therapy. By doing this, the psychologist runs the risk of losing contact with the real world.

Philosophy

Armstrong, D. M. *What is a Law of Nature?* Cambridge: Cambridge University Press, 1975.

Ayer, A. J. *The Problems of Knowledge*. New York: Penguin, 1957.

Berkeley, George. *The Works of George Berkeley, Bishop of Cloyne*. Ed. A. A. Luce and T. E. Jessop. London: Thomas Nelson and Sons, 1948–1957. 9 vols.

*——. *A Treatise Concerning the Principles of Human Knowledge* and *Three Dialogues Between Hylas and Philonous*. J. Dancy, ed. Oxford: Oxford University Press, 1998.

BonJour, Laurence. "Externalist Theories of Empirical Knowledge." *Midwest Studies in Philosophy* 5 (1980): pp. 59–60.

——. *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press, 1985.

Bostrom, Nick. "Are We Living in a Computer Simulation?" *The Philosophical Quarterly* 53 (2003): pp. 243–55. See also: www.simulation-argument.com/simulation.html [Bostom 2003a].

*——. "The Simulation Argument: Why the Probability that You Are Living in a Matrix is Quite High." *Times Higher Education Supplement*, May 16, 2003. See also: www.simulation-argument.com/matrix.html [Bostrom 2003b].

*Chalmers, David. "The Matrix as Metaphysics," 2002, at: http://consc.net/papers/matrix.html

Chisholm, Roderick. *Theory of Knowledge*. 3rd. ed., Englewood Cliffs, NJ: Prentice Hall, 1989.

*Dainton, Barry. "Innocence Lost: Simulation Scenarios: Prospects and Consequences," 2002, at: www.simulation-argument.com/dainton.pdf

Dancy, Jonathan. *Introduction to Contemporary Epistemology*. Oxford: Blackwell, 1985.

Davidson, Donald. "On the Very Idea of a Conceptual Scheme." *Inquiries into Truth and Interpretation*. Oxford University Press, 2001.

Descartes, René. *Meditations on First Philosophy*, I and II (in Chapter 1). *Discourse on Method and Meditations on First Philosophy*, 3rd ed. Trans. Donald Cress. Indianapolis: Hackett Publishing, 1998. Originally published in 1641.

DeRose, Keith and Ted Warfield. *Skepticism: A Contemporary Reader*. Oxford: Oxford University Press, 1999.

Dretske, Fred. *Perception, Knowledge, and Belief*. Cambridge University Press, 2003.
Feldman, Fred. *A Cartesian Introduction to Philosophy*. New York: McGraw Hill, 1986.
Greco, John and Ernest Sosa, eds. *The Blackwell Guide to Epistemology*. Oxford: Blackwell, 1998.
Hegel, G. W. F. *Phenomenology of Mind* (also known as *Phenomenology of Spirit*). Multiple editions in print.
Hirst, R. J. *Perception and the External World*. Macmillan, 1965.
Howson, Colin. *Hume's Problem: Induction and the Justification of Belief*. Oxford University Press, 2003.
Hume, David. *A Treatise of Human Nature*. Multiple editions in print.
——. *Enquiry Concerning Human Understanding*. Multiple editions in print.
Kant, Immanuel. *Critique of Pure Reason*. Multiple editions in print.
Kelley, David. *The Evidence of the Senses: A Realist Theory of Perception*. Louisiana State University Press, 1988.
Kripke, Saul. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980.
Lehrer, Keith. *Theory of Knowledge.* Boulder, CO: Westview Press, 1990.
*Leslie, John. *The End of the World: The Science and Ethics of Human Extinction*. London: Routledge, 1996; revised paperback ed., 1998.
Moore, G. E. *Philosophical Papers*. Allen & Unwin, 1959.
Morick, H. ed. *Wittgenstein and the Problem of Other Minds*. McGraw Hill, 1967.
*Nozick, Robert. *Philosophical Explanations*. Cambridge: Cambridge University Press, 1981.
*Plantinga, Alvin. *Warrant: The Current Debate*. Oxford: Oxford University Press, 1993.
Plato. "The Allegory of the Cave" (from Chapter 2). *Republic*. 2nd ed. Trans. G. M. A. Grube, revision C. D. C. Reeve. Indianapolis: Hackett Publishing, 1992. pp. 186–92. Originally published *c.*380 BCE.
Putnam, Hilary. *Reason, Truth, and History*. Cambridge: Cambridge University Press, 1981.
Reid, Thomas. *An Inquiry into the Human Mind on the Principles of Common Sense*. Derek R. Brookes, ed. University Park: Pennsylvania State University Press, 1997.
Unger, Peter. *Ignorance*. Oxford University Press, 1979.
Weatherson, Brian. "Are You A Sim?" *The Philosophical Quarterly* 53 (2003): pp. 425–31.