

duces the same motion in the brain as when the foot is hurt and, consequently, that the mind feels the same pain "in the foot." And the point holds for other sensations as well.

Finally, I notice that, since only one sensation can be produced by a given motion of the part of the brain that directly affects the mind, the best conceivable sensation for it to produce is the one that is most often useful for the maintenance of the healthy man. Experience teaches, however, that all the sensations put in us by nature *are* of this sort and therefore that there is nothing in our sensations which doesn't testify to God's power and goodness. For example, when the nerves in the foot are moved with unusual violence, the motion is communicated through the middle of the spine to the center of the brain, where it signals the mind to sense a pain "in the foot." This urges the mind to view the pain's cause as harmful to the foot and to do what it can to remove the cause. Of course, God could have so designed man's nature that the same motion of the brain showed the mind something else (such as itself as a motion in the brain, or as a motion in the foot, or as a motion somewhere between the brain and foot), but no alternative to the way things are would be as conducive to the maintenance of the body. Similarly, when we need drink, the throat becomes dry, the dryness moves the nerves of the throat thereby moving the center of the brain, and the brain's movements cause the sensation of thirst in the mind. It's the sensation of thirst that is produced, because no information about our condition is more useful to us than that we need to get something to drink in order to remain healthy. And the same is true in other cases.

This makes it completely obvious that, despite God's immense goodness, the nature of man (whom we now view as a composite of mind and body) cannot fail to be deceptive. For, if something produces the movement usually associated with a bad foot in the nerve running from foot to brain or in the brain itself (rather than in the foot), a pain is felt "in the foot," and the senses are deceived by my nature. Since this motion in the brain must always bring the same sensation to mind, and since the motion's cause more often *is* something hurting the foot than something elsewhere, it is in accordance with reason that the motion always shows the mind a pain in the foot rather than elsewhere. And, if dryness of the throat

arises, not (as usual) from drink's being conducive to the body's health, but (as happens in dropsy) from some other cause, it is much better that we are deceived than that we are generally deceived when our bodies are sound. And so on for other sensations.

In addition to helping me to be aware of the errors to which my nature is subject, these reflections help me readily to correct or avoid those errors. I know that sensory indications of what is good for my body are more often true than false; I can almost always examine a given thing with several senses; and I can also use my memory (which connects the present to the past) and my understanding (which has now ascertained all the causes of error). Hence, I need no longer fear that what the senses daily show me is unreal. I should reject the exaggerated doubts of the past few days as ridiculous. This is especially true of the chief ground for these doubts—namely, that I couldn't distinguish dreaming from being awake—for I now notice that dreaming and being awake are importantly different: the events in dreams are not linked by memory to the rest of my life like those that happen while I am awake. If, while I am awake, someone were suddenly to appear and then immediately to disappear without my seeing where he came from or went to (as happens in dreams), I would justifiably judge that he was not a real man, but a ghost—or, better, an apparition created in my brain. But, if I distinctly observe something's source, its place, and the time at which I learn about it, and if I grasp an unbroken connection between it and the rest of my life, I am quite sure that it is something in my waking life rather than in a dream. And I ought not to have the slightest doubt about the reality of such things, if I have examined them with all my senses, my memory, and my understanding without finding any conflicting evidence. For, from the fact that God is not a deceiver, it follows that I am not deceived in any case of this sort. Since the need to act does not always allow time for such a careful examination, however, we must admit the likelihood of men's erring about particular things and acknowledge the weakness of our nature.

STUDY QUESTIONS

1. Descartes comes to the conclusion that "there are no reliable signs by which I can distinguish sleeping

- from waking." What considerations lead him to this conclusion?
2. Why does Descartes suppose, at the end of Meditation I, that there is an evil demon "who works as hard as he can to deceive me"?
3. Why does Descartes think he knows for certain that he exists? Is he entitled to this conclusion?

4. In Meditation III, Descartes argues that his idea of God could not have come from him, and so God must exist. How does this argument go?
5. Descartes argues in Meditation VI that he can exist without his body. What is his argument for this? Can you similarly imagine existing without a body?

Bad Dreams, Evil Demons, and the Experience Machine: Philosophy and *The Matrix*

CHRISTOPHER GRAU

Christopher Grau (1970–) is professor of philosophy at Florida International University. He studied at Johns Hopkins University under Susan Wolf and works on ethics, metaphysics, and philosophy and film. He edited and contributed to the volume *Philosophers Explore the Matrix*.

I. Dream Skepticism

MORPHEUS: Have you ever had a dream, Neo, that you were so sure was real?

MORPHEUS: What if you were unable to wake from that dream, Neo? How would you know the difference between the dreamworld and the real world?

Neo has woken up from a hell of a dream—the dream that was his life. How was he to know? The cliché is that if you are dreaming and you pinch yourself, you will wake up. Unfortunately, things aren't quite that simple. It is the nature of most dreams that we take them for reality—while dreaming, we are unaware that we are in a dream world.

Of course, we eventually wake up, and when we do, we realize that our experience was all in our mind. Neo's predicament makes us wonder, though: how can any of us be sure that we have ever *genuinely* woken up? Perhaps, like Neo prior to his downing the red pill, our dreams thus far have in fact been dreams *within* a dream.

The idea that what we take to be the real world could all be just a dream is familiar to many students of philosophy, poetry, and literature. Most of us, at one time or another, have been struck with the thought that we might mistake a dream for reality, or reality for a dream. Arguably the most famous exponent of this worry in the Western philosophical tradition is the seventeenth-century French philosopher René Descartes. In an attempt to provide a firm foundation for knowledge, he began his *Meditations* by clearing the philosophical ground through doubting all that could be doubted. This was done, in part, in order to determine if anything that could count as certain knowledge could survive such rigorous and systematic skepticism. Descartes takes the first step

From "Bad Dreams, Evil Demons, and the Experience Machine: Philosophy and *The Matrix*," in Christopher Grau, ed., *Philosophers Explore the Matrix*, pp. 10–23. Copyright © 2005 by Oxford University Press. Reprinted by permission of the publisher.

toward this goal by raising (through his fictional narrator) the possibility that we might be dreaming:

How often, asleep at night, am I convinced of just such familiar events—that I am here in my dressing-gown, sitting by the fire—when in fact I am lying undressed in bed! Yet at the moment my eyes are certainly wide awake when I look at this piece of paper; I shake my head and it is not asleep; as I stretch out and feel my hand I do so deliberately, and I know what I am doing. All this would not happen with such distinctness to someone asleep. Indeed! As if I did not remember other occasions when I have been tricked by exactly similar thoughts while asleep! As I think about this more carefully, I see plainly that there are never any sure signs by means of which being awake can be distinguished from being asleep. The result is that I begin to feel dazed, and this very feeling only reinforces the notion that I may be asleep.¹

When we dream, we are often blissfully ignorant that we are dreaming. Given this, and the fact that dreams often seem as vivid and “realistic” as real life, how can you rule out the possibility that you might be dreaming even now, as you sit and read this? This is the kind of perplexing thought that Descartes forces us to confront. It seems we have no justification for the belief that we are not dreaming. If so, then it seems we similarly have no justification in thinking that the world we experience is the *real* world. Indeed, it becomes questionable whether we are justified in thinking that *any* of our beliefs are true.

The narrator of Descartes’s *Meditations* worries about this, but he ultimately maintains that the possibility that one might be dreaming cannot by itself cast doubt on all we think we know; he points out that even if all our sensory experience is but a dream, we can still conclude that we have *some* knowledge of the nature of reality. Just as a painter cannot create *ex nihilo* but must rely on pigments with which to create her image, certain elements of our thought must exist prior to our imaginings. Among the items of knowledge that Descartes thought survived dream skepticism are truths arrived at through the use of reason, such as the truths of mathematics: “For whether I am awake or asleep, two and three

added together are five, and a square has no more than four sides” (296–297).

While such an insight offers little comfort to someone wondering whether the people and objects she confronts are genuine, it served Descartes’s larger philosophical project: he sought, among other things, to provide a foundation for knowledge in which truths arrived at through reason are given priority over knowledge gained from experience. (This bias shouldn’t surprise those who remember that Descartes was a brilliant mathematician in addition to being a philosopher.) Descartes was not himself a skeptic—he employs this skeptical argument so as to help remind the reader that the truths of mathematics (and other truths of reason) are on firmer ground than the data provided to us by our senses.

Despite the fact that Descartes’s ultimate goal was to demonstrate how genuine knowledge is possible, he proceeds in *The Meditations* to utilize a much more radical skeptical argument, one that casts doubt on even his beloved mathematical truths. In the next section we will see that, many years before the Wachowskis dreamed up *The Matrix*, Descartes had imagined an equally terrifying possibility.

II. Brains in Vats and the Evil Demon

MORPHEUS: What is the Matrix? Control.

MORPHEUS: The Matrix is a computer-generated dreamworld built to keep us under control in order to change a human being into this. [holds up a battery]

NEO: No! I don’t believe it! It’s not possible!

Before breaking out of the Matrix, Neo’s life was not what he thought it was. It was a lie. Morpheus describes it as a “dreamworld,” but unlike a dream, this world is not the creation of Neo’s mind. The truth is more sinister: the world is a creation of the artificially intelligent computers that have taken over the Earth and have subjugated humanity in the process. These creatures have fed Neo a simulation that he couldn’t possibly help but take as the real thing. What’s worse, it isn’t clear how any of us can know with certainty that we are not in a position similar to Neo before his “rebirth.” Our ordinary

confidence in our ability to reason and our natural tendency to trust the deliverances of our senses can both come to seem rather naive once we confront this possibility of deception. A viewer of *The Matrix* is naturally led to wonder: how do I know I am not in the Matrix? How do I know for sure that my world is not also a sophisticated charade, put forward by some superhuman intelligence in such a way that I cannot possibly detect the ruse? Descartes suggested a similar worry: the frightening possibility that all of one’s experiences might be the result of a powerful outside force, a “malicious demon”:

And yet firmly 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 about 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? What is more, just as I consider that others sometimes go astray in cases where they think they have the most perfect knowledge, how do I know that God has not brought it about that I too go wrong every time I add two and three or count the sides of a square, or in some even simpler matter, if that is imaginable? But perhaps God would not have allowed me to be deceived in this way, since he is said to be supremely good; . . . I will suppose therefore that not God, who is supremely good and the source of truth, but rather some malicious demon of the utmost power and cunning has employed all his energies in order to deceive me. I shall think that the sky, the air, the earth, colours, shapes, sounds and all external things are merely the delusions of dreams which he has devised to ensnare my judgment. (297–298)

The narrator of Descartes’s *Meditations* concludes that none of his former opinions are safe. Such a demon could not only deceive him about his perceptions, it could conceivably cause him to go wrong when performing even the simplest acts of reasoning.

This radical worry seems inescapable. How could you possibly prove to yourself that you are not in the kind of nightmarish situation Descartes describes? It would seem that any argument, evidence, or proof you might put forward could easily be yet

another trick played by the demon. As ludicrous as the idea of the evil demon may sound at first, it is hard, upon reflection, not to share Descartes’s worry: for all you know, you may well be a mere plaything of such a malevolent intelligence. More to the point of our general discussion: for all you know, you may well be trapped in the Matrix.

Many contemporary philosophers have discussed a similar skeptical dilemma that is a bit closer to the scenario described in *The Matrix*. It has come to be known as the “brain in a vat” hypothesis, and one powerful formulation of the idea is presented by philosopher Jonathan Dancy:

You do not know that you are not a brain, suspended in a vat full of liquid in a laboratory, and wired to a computer which is feeding you your current experiences under the control of some ingenious technician scientist (benevolent or malevolent according to taste). For if you were such a brain, then, provided that the scientist is successful, nothing in your experience could possibly reveal that you were; for your experience is *ex hypothesi* identical with that of something which is not a brain in a vat. Since you have only your own experience to appeal to, and that experience is the same in either situation, nothing can reveal to you which situation is the actual one.²

If you cannot know whether you are in the real world or in the world of a computer simulation, you cannot be sure that your beliefs about the world are true. And, what was even more frightening to Descartes, in this kind of scenario it seems that your ability to reason is no safer than the deliverances of the senses: the evil demon or malicious scientist could be ensuring that your reasoning is just as flawed as your perceptions. As you have probably already guessed, there is no easy way out of this philosophical problem (or at least there is no easy *philosophical* way out). Philosophers have proposed a dizzying variety of “solutions” to this kind of skepticism but, as with many philosophical problems, there is nothing close to unanimous agreement regarding how the puzzle should be solved.

Descartes’s own way out of his evil-demon skepticism was to first argue that one cannot genuinely doubt the existence of oneself. He pointed out that

all thinking presupposes a thinker: even in doubting, you realize that there must at least be a self which is doing the doubting. (Thus Descartes's most famous line: "I think, therefore I am.") He then went on to claim that, in addition to our innate idea of self, each of us has an idea of God as an all-powerful, all-good, and infinite being implanted in our minds and that this idea could only have come from God. Since this shows us that an all-good God does exist, we can have confidence that he would not allow us to be so drastically deceived about the nature of our perceptions and their relationship to reality. While Descartes's argument for the existence of the self has been tremendously influential and is still actively debated, few philosophers have followed him in accepting his particular theistic solution to skepticism about the external world.

One of the more interesting contemporary challenges to the kind of radical skepticism suggested by Descartes has come from philosopher Hilary Putnam. His point is not so much to defend our ordinary claims to knowledge as to question whether the brain-in-a-vat hypothesis is coherent, given certain plausible assumptions about how our language refers to objects in the world. He asks us to consider a variation on the standard brain-in-a-vat story that is uncannily similar to the situation described in *The Matrix*:

Instead of having just one brain in a vat, we could imagine that all human beings (perhaps all sentient beings) are brains in a vat (or nervous systems in a vat in case some beings with just nervous systems count as "sentient"). Of course, the evil scientist would have to be outside? or would he? Perhaps there is no evil scientist, perhaps (though this is absurd) the universe just happens to consist of automatic machinery tending a vat full of brains and nervous systems. This time let us suppose that the automatic machinery is programmed to give us all a *collective* hallucination, rather than a number of separate unrelated hallucinations. Thus, when I seem to myself to be talking to you, you seem to yourself to be hearing my words. . . . I want now to ask a question which will seem very silly and obvious (at least to some people, including some very sophisticated philosophers), but which will take us to real philosophical depths rather quickly. Suppose this whole story were actu-

ally true. Could we, if we were brains in a vat in this way, *say* or *think* that we were?³

Putnam's surprising answer is that we cannot coherently think that we are brains in vats, and so skepticism of that kind can never really get off the ground. While it is difficult to do justice to Putnam's ingenious argument in a short summary, his point is roughly as follows: not everything that goes through our heads is a genuine thought, and far from everything we say is a meaningful utterance. Sometimes we get confused or think in an incoherent manner; sometimes we say things that are simply nonsense. Of course, we don't always realize at the time that we aren't making sense; sometimes we earnestly believe we are saying (or thinking) something meaningful. High on nitrous oxide, the philosopher William James was convinced he was having profound insights into the nature of reality; he was convinced that his thoughts were both sensical and important. Upon sobering up and looking at the notebook in which he had written his drug-addled thoughts, he saw only gibberish.

Just as I might say a sentence that is nonsense, I might also use a name or a general term that is meaningless in the sense that it fails to hook up to the world. Philosophers talk of such a term as "failing to refer" to an object. In order to successfully refer when we use language, there must be an appropriate relationship between the speaker and the object referred to. If a dog playing on the beach manages to scrawl the word "Ed" in the sand with a stick, few would want to claim that the dog actually meant to refer to someone named Ed. Presumably the dog doesn't know anyone named Ed, and even if he did, he wouldn't be capable of intending to write Ed's name in the sand. The point of such an example is that words do not refer to objects "magically" or intrinsically: certain conditions must be met in the world in order for us to accept that a given written or spoken word has any meaning and whether it actually refers to anything at all.

Putnam claims that one condition which is crucial for successful reference is that there be an appropriate causal connection between the object referred to and the speaker referring. Specifying exactly what should count as "appropriate" here is a notoriously difficult task, but we can get some idea

of the kind of thing required by considering cases in which reference fails through an inappropriate connection: if someone unfamiliar with the film *The Matrix* manages to blurt out the word "Neo" while sneezing, few would be inclined to think that this person has actually *referred* to the character Neo. The kind of causal connection between the speaker and the object referred to (Neo) is just not in place. For reference to succeed, it can't be simply accidental that the name was uttered. (Another way to think about it: the sneezer would have uttered "Neo" even if the film *The Matrix* had never been made.)

The difficulty, according to Putnam, in coherently supposing the brain-in-a-vat story to be true is that brains raised in such an environment could not successfully refer to genuine brains, or vats, or anything else in the real world. Consider the example of some people who lived their entire lives in the Matrix: when they talk of "chickens," they don't actually refer to real chickens; at best they refer to the computer representations of chickens that have been sent to their brains. Similarly, when they talk of human bodies being trapped in pods and fed data by the Matrix, they don't successfully refer to real bodies or pods. They can't refer to physical bodies in the real world because they cannot have the appropriate causal connection to such objects. Thus, if someone were to utter the sentence "I am simply a body stuck in a pod somewhere being fed sensory information by a computer," that sentence would itself be necessarily false. If the person is in fact not trapped in the Matrix, then the sentence is straightforwardly false. If the person is trapped in the Matrix, then he can't successfully refer to real human bodies when he utters the term "human body," and so it appears that under this circumstance, his statement must *also* be false. Such a person seems thus doubly trapped: incapable of knowing that he is in the Matrix and even incapable of successfully expressing the thought that he might be in the Matrix! (Could this be why at one point Morpheus tells Neo that "no one can be told what the Matrix is"?)

Putnam's argument is controversial, but it is noteworthy because it shows that the kind of situation described in *The Matrix* raises not just the expected philosophical issues about knowledge and skepticism, but more general issues regarding meaning, language, and the relationship between the mind and the world.

III. The Value of Reality: Cypher and the Experience Machine

CYPHER: You know, I know that this steak doesn't exist. I know when I put it in my mouth, the Matrix is telling my brain that it is juicy and delicious. After nine years, do you know what I've realized?

CYPHER: Ignorance is bliss.

AGENT SMITH: Then we have a deal?

CYPHER: I don't want to remember nothing. Nothing! You understand? And I want to be rich. Someone important. Like an actor. You can do that, right?

AGENT SMITH: Whatever you want, Mr. Reagan.

Cypher is not a nice guy, but is he an unreasonable guy? Is he right to want to get re-inserted into the Matrix? Many want to say no, but giving reasons for why his choice is a bad one is not an easy task. After all, so long as his experiences will be pleasant, how can his situation be worse than the inevitably crappy life he would lead outside of the Matrix? What could matter beyond the quality of his experience? Remember, once he's back in, living his fantasy life, he won't even know he made the deal. What he doesn't know can't hurt him, right?

Is feeling good the only thing that has value in itself? The question of whether only conscious experience can ultimately matter is one that has been explored in depth by several contemporary philosophers. In the course of discussing this issue in his 1974 book *Anarchy, State, and Utopia*, Robert Nozick introduced a thought-experiment that has become a staple of introductory philosophy classes everywhere. It is known as "the experience machine":

Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? . . . Of course, while in the tank you won't know that you're there; you'll think it's all actually happening.

Others can also plug in to have the experiences they want, so there's no need to stay unplugged to serve them. (Ignore problems such as who will service the machines if everyone plugs in.) Would you plug in? *What else can matter to us, other than how our lives feel from the inside?*⁴

Nozick goes on to argue that other things do matter to us: for instance, that we actually do certain things, as opposed to simply having the experience of doing them. Also, he points out that we value being (and becoming) certain kinds of people. I don't just want to have the experience of being a decent person, I want to actually *be* a decent person. Finally, Nozick argues that we value contact with reality in itself, independent of any benefits such contact may bring through pleasant experience: we want to know we are experiencing the real thing. In sum, Nozick thinks that it matters to most of us, often in a rather deep way, that we be the authors of our lives and that our lives involve interacting with the world, and he thinks that the fact that most people would not choose to enter into such an experience machine demonstrates that they do value these other things. As he puts it: "We learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it" (311).

While Nozick's description of his machine is vague, it appears that there is at least one important difference between it and the simulated world of *The Matrix*. Nozick implies that people hooked up to the experience machine will not be able to exercise their agency—they become the passive recipients of pre-programmed experiences. This apparent loss of free will is disturbing to many people, and it might be distorting people's reactions to the case and clouding the issue of whether they value contact with reality *per se*. The Matrix seems to be set up in such a way that one can enter it and retain one's free will and capacity for decision making, and perhaps this makes it a significantly more attractive option than the experience machine that Nozick describes.

Nonetheless, a loss of freedom is not the only disturbing aspect of Nozick's story. As he points out, we seem to mourn the loss of contact with the real world as well. Even if a modified experience machine is pre-

sented to us, one which allows us to keep our free will but enter into an entirely virtual world, many would still object that permanently going into such a machine involves the loss of something valuable.

Cypher and his philosophical comrades are likely to be unmoved by such observations. So what if most people are hung-up on "reality" and would turn down the offer to permanently enter an experience machine? Most people might be wrong. All their responses might show is that such people are superstitious, or irrational, or otherwise confused. Maybe they think something could go wrong with the machines, or maybe they keep forgetting that while in the machine they will no longer be aware of their choice to enter the machine. Perhaps those hesitant to plug in don't realize that they value being active in the real world only because normally that is the most reliable way for them to acquire the pleasant experience that they value in itself. In other words, perhaps our free will and our capacity to interact with reality are means to a further end; they matter to us because they allow us access to what really matters: pleasant conscious experience. To think the reverse, that reality and freedom have value in themselves (what philosophers sometimes call *non-derivative* or *intrinsic value*), is simply to put the cart before the horse. After all, Cypher could reply, what would be so great about the capacity to freely make decisions or the ability to be in the real world if neither of these things allowed us to feel good?

Peter Unger has taken on these kinds of objections in his discussion of "experience inducers." He acknowledges that there is a strong temptation when in a certain frame of mind to agree with this kind of Cypheresque reasoning, but he argues that this is a temptation we ought to try to resist. Cypher's vision of value is too easy and too simplistic. We are inclined to think that only conscious experience can really matter in part because we fall into the grip of a particular picture of what values *must* be like, and this in turn leads us to stop paying attention to our actual values. We make ourselves blind to the subtlety and complexity of our values, and we then find it hard to understand how something that doesn't affect our consciousness could sensibly matter to us. If we stop and reflect on what we really do care about, however, we come across

some surprisingly everyday examples that don't sit easily with Cypher's claims:

Consider life insurance. To be sure, some among the insured may strongly believe that, if they die before their dependents do, they will still observe their beloved dependents, perhaps from a heaven on high. But others among the insured have no significant belief to that effect. . . . Still, we all pay our premiums. In my case, this is because, even if I will never experience anything that happens to them, I still want things to go better, rather than worse, for my dependents. No doubt, I am rational in having this concern.⁵

As Unger goes on to point out, it seems contrived to chalk up all examples of people purchasing life insurance to cases in which someone is simply trying to benefit (while alive) from the favorable impression such a purchase might make on the dependents. In many cases it seems ludicrous to deny that "what motivates us, of course, is our great concern for our dependents' future, whether we experience their future or not" (302). This is not a proof that such concern is rational, but it does show that incidents in which we intrinsically value things other than our own conscious experience might be more widespread than we are at first liable to think. (Other examples include the value we place on not being deceived or lied to—the importance of this value doesn't seem to be completely exhausted by our concern that we might one day become aware of the lies and deception.)

Most of us care about a lot of things independently of the experiences that those things provide for us. The realization that we value things other than pleasant conscious experience should lead us to at least wonder if the legitimacy of this kind of value hasn't been too hastily dismissed by Cypher and his ilk. After all, once we see how widespread and commonplace our other nonderivative concerns are, the insistence that conscious experience is the *only* thing that has value in itself can come to seem downright peculiar. If purchasing life insurance seems like a rational thing to do, why shouldn't the desire to experience reality (rather than some illusory simulation) be similarly rational? Perhaps the best test of the ra-

tionality of our most basic values is actually whether they, taken together, form a consistent and coherent network of attachments and concerns. (Do they make sense in light of each other and in light of our beliefs about the world and ourselves?) It isn't obvious that valuing interaction with the real world fails this kind of test.

Of course, pointing out that the value I place on living in the real world coheres well with my other values and beliefs will not quiet the defender of Cypher, as he will be quick to respond that the fact that my values all cohere doesn't show that they are all justified. Maybe I hold a bunch of exquisitely consistent but thoroughly irrational values.

The quest for some further justification of my basic values might be misguided, however. Explanations have to come to an end somewhere, as Ludwig Wittgenstein once famously remarked. Maybe the right response to a demand for justification here is to point out that the same demand can be made to Cypher: "Just what justifies your exclusive concern with pleasant conscious experience?" It seems as though nothing does—if such concern is justified, it must be somehow self-justifying, but if that is possible, why shouldn't our concerns for other people and our desire to live in the real world also be self-justifying? If those can also be self-justifying, then maybe what we don't experience should matter to us, and perhaps what we don't know *can* hurt us. . . .

Further Reading

Those seeking to go further should certainly begin with the rest of Descartes's *Meditations*. Currently, the best edition is *The Philosophical Writings of Descartes*, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge University Press 1984). A solid and comprehensive introduction to epistemology is Jonathan Dancy's *Introduction to Contemporary Epistemology* (Blackwell 1985). For slightly more advanced treatments, I recommend Barry Stroud's *The Significance of Philosophical Scepticism* (Oxford University Press 1984) and P. F. Strawson's *Skepticism and Naturalism: Some Varieties* (Columbia University Press 1983). On the question of the justification of values, my comments here

draw on the insights of Mark Johnston in "Reasons and Reductionism" (*Philosophical Review* 1992), Thomas Nagel's essay "Death" (*Nous* 1970), and Peter Unger's *Identity, Consciousness, and Value* (Oxford University Press 1990).

NOTES

1. René Descartes, *Meditations on First Philosophy*, ed. and trans. J. Cottingham (Cambridge University Press 1985); see appendix, pp. 295–296. Further citations will be in the text.
2. Jonathan Dancy, *Introduction to Contemporary Epistemology* (Blackwell 1985), p. 10.
3. Hilary Putnam, *Reason, Truth, and History* (Cambridge University Press 1981); see appendix, pp. 317–318.
4. Robert Nozick, *Anarchy, State, and Utopia* (Basic 1974); see appendix, p. 310. Further citations will be to this appendix and appear in the text.

Excerpt from *Philosophical Explanations*

ROBERT NOZICK

Robert Nozick (1938–2002) was an American philosopher and professor at Harvard who contributed influential ideas to many areas of philosophy, epistemology, and political philosophy, in particular. His works include *Philosophical Explanations*, *The Examined Life*, and *Anarchy, State, and Utopia*.

I. Knowledge

Conditions for Knowledge

Our task is to formulate further conditions to go alongside

Reprinted by permission of the publisher from *Philosophical Explanations* by Robert Nozick, pp. 172–178; and 204–211, Cambridge, Mass: The Belknap Press of Harvard University Press. Copyright © 1981 by the President and Fellows of Harvard College.

5. Peter Unger, *Identity, Consciousness, and Value* (Oxford University Press 1990), p. 301. Further citations will be in the text.

STUDY QUESTIONS

1. Should we take seriously the hypothesis that we might be in the Matrix? Why or why not? Would finding out that you were in the Matrix change the way you live your life?
2. Why does Putnam think that if we were brains in vats, we couldn't say or think that we were? How important is this conclusion?
3. Would you get into Nozick's Experience Machine? Why or why not? Do you think there is any value at all to being "in touch" with reality?
4. What is Unger's life insurance analogy supposed to show? Is the analogy successful?

- (1) p is true
- (2) S believes that p .

We would like each condition to be necessary for knowledge, so any case that fails to satisfy it will not be an instance of knowledge. Furthermore, we would like the conditions to be jointly sufficient for knowledge, so any case that satisfies all of them will be an instance of knowledge. We first shall formulate conditions that seem to handle ordinary cases correctly, classifying as knowledge cases which

are knowledge, and as nonknowledge cases which are not; then we shall check to see how these conditions handle some difficult cases discussed in the literature.

The causal condition on knowledge, previously mentioned, provides an inhospitable environment for mathematical and ethical knowledge; also there are well-known difficulties in specifying the type of causal connection. If someone floating in a tank oblivious to everything around him is given (by direct electrical and chemical stimulation of the brain) the belief that he is floating in a tank with his brain being stimulated, then even though that fact is part of the cause of his belief, still he does not know that it is true.

Let us consider a different third condition:

- (3) If p weren't true, S wouldn't believe that p .

Throughout this work, let us write the subjunctive 'if-then' by an arrow, and the negation of a sentence by prefacing "not-" to it. The above condition thus is rewritten as:

- (3) $\text{not-}p \rightarrow \text{not-}(S \text{ believes that } p)$.

This subjunctive condition is not unrelated to the causal condition. Often when the fact that p (partially) causes someone to believe that p , the fact also will be causally necessary for his having the belief—without the cause, the effect would not occur. In that case, the subjunctive condition 3 also will be satisfied. Yet this condition is not equivalent to the causal condition. For the causal condition will be satisfied in cases of causal overdetermination, where either two sufficient causes of the effect actually operate, or a back-up cause (of the same effect) would operate if the first one didn't; whereas the subjunctive condition need not hold for these cases. When the two conditions do agree, causality indicates knowledge because it acts in a manner that makes the subjunctive 3 true.

The subjunctive condition 3 serves to exclude cases of the sort first described by Edward Gettier, such as the following. Two other people are in my office and I am justified on the basis of much evidence in believing the first owns a Ford car; though he (now) does not, the second person (a stranger to

me) owns one. I believe truly and justifiably that someone (or other) in my office owns a Ford car, but I do not know someone does. Concluded Gettier, knowledge is not simply justified true belief.

The following subjunctive, which specifies condition 3 for this Gettier case, is not satisfied: if no one in my office owned a Ford car, I wouldn't believe that someone did. The situation that would obtain if no one in my office owned a Ford is one where the stranger does not (or where he is not in the office); and in that situation I still would believe, as before, that someone in my office does own a Ford, namely, the first person. So the subjunctive condition 3 excludes this Gettier case as a case of knowledge.

The subjunctive condition is powerful and intuitive, not so easy to satisfy, yet not so powerful as to rule out everything as an instance of knowledge. A subjunctive conditional "if p were true, q would be true", $p \rightarrow q$, does not say that p entails q or that it is logically impossible that p yet not- q . It says that in the situation that would obtain if p were true, q also would be true. This point is brought out especially clearly in recent 'possible-worlds' accounts of subjunctives: the subjunctive is true when (roughly) in all those worlds in which p holds true that are closest to the actual world, q also is true. (Examine those worlds in which p holds true closest to the actual world, and see if q holds true in all these.) Whether or not q is true in p worlds that are still farther away from the actual world is irrelevant to the truth of the subjunctive. I do not mean to endorse any particular possible-worlds account of subjunctives, nor am I committed to this type of account. I sometimes shall use it, though, when it illustrates points in an especially clear way.*

* If the possible-worlds formalism is used to represent counterfactuals and subjunctives, the relevant worlds are not those p worlds that are closest or most similar to the actual world, unless the measure of closeness or similarity is: what would obtain if p were true. Clearly, this cannot be used to explain when subjunctives hold true, but it can be used to represent them. Compare utility theory which represents preferences but does not explain them. Still, it is not a trivial fact that preferences are so structured that they can be represented by a real-valued function, unique up to a positive linear transformation, even though the representation (by itself) does not explain these preferences. Similarly, it would be of