KNOWING ONE'S MIND

Williams College Campus Lecture, 8 February 2007

Joe Cruz, Department of Philosophy and Program in Cognitive Science

In one of the more compelling introductions to philosophy, Bertrand Russell begins with this question: "Is there any knowledge in the world that is so certain that no reasonable man could doubt it?" (Presumably he means to include women.) "So certain that no reasonable man could doubt it." And it's a good question to begin an introduction to philosophy with, because so often, philosophy is in the mode of skepticism, so often it's in the mode of offering a critical assessment of conventional wisdom. So, Russell wonders, is there anything so certain that no reasonable man could doubt it. And when we embark on this question, I suppose we have to ask about the question itself, we have to wonder what Russell's talking about—right? We have to wonder what certainty is.

So, what is certainty? It can't merely be powerful confidence, it can't merely be something like the assurance that we feel for ordinary knowledge claims. After all, there are lots of things that I *know*: I know that two plus two is four, I know that water is H_2O, I know that I'm standing here before you. But I'd balk if you pressed me and asked me whether I was *certain* about these things—well, I don't know if I'm certain about these things, I believe them on what I take is good evidence, I have a considerable confidence in these claims, I'd even bet a whole lot on at least some of them, but *certain* about it? I'm not sure about that. So we have to ask: what more is required than our confidence, then something like reasonable belief on plausible evidence? What's certainty?

Well, here's a proposal—here's a first pass—and here's how I'll treat certainty (sneaking up on the question of "knowing one's mind," and it won't surprise you to hear that I'm going to wonder whether we know anything for certain about our own minds—so, sneaking up on that question, here's a first pass on certainty). Certainty is, well, that condition when the claim you have before you can't be false, it's *inconceivable* that it's false. So, I'm going to claim that a claim is certain if it's inconceivable for it to be false. And that's different, I say—that's different from a claim being necessarily true. Certainty is what philosophers will call an epistemic quantity—having to do with knowledge, having to do with our conception of the world. Certainty is an epistemic quality, as opposed to *necessity*—something's being necessarily true—which philosophers will identify as a metaphysical quantity. So, it might well be—it might well be that some very complicated proof in mathematics yields a necessary truth at the end. It might be that the proof is an excellent one, but it's going to take some work on the part of the mathematician to reveal the necessity of that conclusion, and that goes to show that if she's *certain* about the conclusion, that's a different achievement than the necessity itself (certainty being an epistemic quantity, necessity being a metaphysical quantity).

So, this puts us in the ballpark: certainty is going to have to do with the inconceivability of something's being false, and that's different from the mere metaphysical quantity of it's being necessarily true. This suggests that certainty is a kind of reflexive or conceptual achievement or state that you might be in —a state of certainty—not merely psychological confidence, but something beyond psychological confidence, the state of being so sure that it's inconceivable that you're wrong. Are there any plausible candidates, you're wondering, any plausible candidates for certainty? Surely not the things I mentioned before, surely not the claim that water is H_2O or two plus to equals four or that I'm standing here before you. Is it conceivable that

those are false? Sure it's conceivable that those are false. We can imagine wild cases where I'm totally wrong about those things, though I had a confidence in them. But are there any plausible candidates for real certainty? Well, how about this: Descartes offers—Second Meditation—Descartes offers this: "I am, I exist, this is certain. How often? As often as I think." How often? As often as I think. This is certain, claims Descartes in the Second Meditation. And you can kinda see what's compelling about this. If there's anything that counts as a firm result in philosophy (we don't get results in philosophy, but every few hundred years something is counted as close to a firm result in philosophy, and this is often thought to be one of them—not such a bad argument, or not such a bad placeholder for an argument). Descartes claims that he can't be wrong about his own existence in the moment that he thinks it, he can't be wrong about his own existence because, well, if he didn't exist, who would there be to think it? How could he be wrong about this as long as he's thinking? "I am, I exist"—that's our candidate for certainty.

Now I don't have to be coy. By the end of the talk, I'm going to argue that Descartes is wrong about this, that he can't be certain that he exists at the moment that he thinks it. I hear—though they won't tell it to me to my face—I hear that in Harper House they think I'm mad (Harper House is where the philosophers have their offices). Maybe that's a mad thing to think, but I'm going to give you the argument over the next 25 or so minutes. I'm going to deny this claim. But here's our first candidate for something's being certain. What does Descartes say again? He says that he can be certain that he exists—"I exist, I am." How often can he be certain? As often as he thinks . . . as often as he thinks. So let's ask: what does he mean by thinking? What is thinking?

Well here's a proposal, a first pass, if you will: thinking as having explicit ideas. And there's some evidence for Descartes thinking of it this way in the Meditations, when he says,

"What is a thinking thing? It's a thing that doubts, that understands, conceives, affirms, denies, wills, refuses." There's this view about activity, sort of a mental activity, that is directed at some particular claim. Something is affirmed, something is doubted, something is willed. So, one sense of thinking, one conception of thinking, that Descartes might be working with in his argument, is thinking as having explicit ideas. So here's thinking as having an explicit idea: the cat is on the mat. I'm now thinking, "The cat is on the mat." (If you're paying attention, you are too.) The cat is on the mat. Not an image of a cat on a mat, not an experience or perception of a cat on a mat, but rather the thought, "The cat is on the mat." So don't get confused: this proposal—of thinking as having explicit ideas—is a proposal that there's something articulated, something conceptual, something explicit, held in the mind, that's different from the play of images or the play of consciousness that unfolds in our normal waking lives. Thinking as having explicit ideas. So "The cat is on the mat," not the conscious experience. And we might say, with Descartes, that this is thinking of thinking as pure intellection—as purely a thought, pure intellection.

Here are some other examples. Here's a question for you: does utilitarianism lead to injustice? (You can work that out after the talk.) There's "I want chocolate," "the sunset is lovely." So, here are some articulated explicit thoughts—having explicit ideas. So, here's another way—yet another way of putting it, to put you in the right neighborhood with this way of thinking: having an idea the content of which is explicitly articulated, clear in your mind, and there is some attitude toward that content, you either want it, or you affirm it, or you wish it, or you deny it, but the key is that there is an explicit content there. Having explicit ideas—this is one way of understanding thinking. And so, we get to ask if Descartes is right to claim that we

can be certain of our own existence at the moment when we're thinking it, or when we're thinking at all—when we're thinking, in this sense.

So, Descartes, in his clever and polite way, offers us an argument for pure intellection. He claims that there has to be this capacity, he claims that there has to be this state that we can be in involving pure intellection. And his argument—those of you who are taking 102 right now will get to it soon enough—his argument can be found in the Sixth Meditation, where he contrasts triangles with chiliogons. And the delightful thing about triangles (chiliogons—thousand-sided figures), the delightful thing about triangles is that you can do all kinds of geometric manipulations on them, you can treat them as mathematical objects, but you can also understand them imagistically or within consciousness, within your mind's eye you can maintain a triangle. That's the delightful thing about triangles. Not so with chiliogons, at least for most people. For most people, while we might still end up being able to perform mathematical manipulations on chiliogons, maintaining a chiliogon in the imagination, having an image of it before our mind's eye, is rather more difficult. And you can check for yourself by imagining a thousand-sided figure (there you are, imagine a thousand-sided figure), and then imagining a figure that has 999 sides, and see if you can discern a difference. Does it feel different to you? Chances are, it does not. And Descartes tries to leverage this into the argument that there must be something like pure intellection, there must be the capacity for having explicit and articulated thoughts that don't have anything to do with conscious experience or have anything to do with imagination. And this is the sense of thinking that we're talking about. So, thinking is having explicit ideas, in this idea.

Question: can you be wrong about the explicit ideas you're having? So, try it out. I invite you to think, "The cat is on the mat." Everybody with me, but for the sleepers? Great. Could you

be wrong that that's what you're thinking right now? Could you be wrong. No, really—could you be wrong? Hard to see you could be wrong—after all, you are the world's expert on, if nothing else, your thoughts. Hard to see how you could be wrong. You're thinking it—"The cat is on the mat" firmly in mind. Maybe you could be certain about that. I wonder. Huh. Here's a Cartesian way of asking the question: could an evil genius, could an evil genius . . . deceive you with respect to your explicit ideas? And this is a kind of metaphorical placeholder for Descartes, a kind of placeholder for this idea of imagining possibilities, imagining the possibility of you being wrong—or, to put it in the way that I earlier did, by engaging in this conceivability task, can you conceive of yourself being wrong about the explicit ideas that you have? Could an evil genius deceive you?

Well let's see, could you be wrong. Look, of course you can't be certain that the cat is on the mat, right? I mean, after all, an evil genius *could* be deceiving you with respect to "the cat is on the mat," or you might be in the Matrix, or you might be dreaming, or you might just be plain wrong—it's a small dog, not a cat. So, you can be wrong about the cat being on the mat, there's no doubt about that. But how about this: I'm certain that I am now thinking, "The cat is on the mat." So, can you be wrong about this claim? I'm certain that I am now thinking, "The cat is on the mat." Well not only can't you be certain of it, it's not even true that that's what you're now thinking. Try it out. Consider the claim, "I'm certain that I am now thinking, 'The cat is on the mat.'" Well as you reflect on that claim—I am now certain that the cat is on the mat—you're not thinking that the cat is on the mat, you're thinking this claim—"I am now certain that I am thinking, 'The cat is on the mat.'" You're reflecting on your certainty, you're engaged in a different task than the task that you set out for yourself. So you started out with this challenge—

“The cat is on the mat,” can I be certain of that? Huh. Well then, I wonder, can I be certain that the cat is on the mat—a different thought.

Ok, how about this one, then—maybe this is a better stab at certainty: “I am now certain I am thinking, ‘The cat is on the mat.’” So I’m reflecting on myself, I’m wondering about myself. I am now certain I am thinking, “The cat is on the mat.” But not only is this not certain, it’s not even true, because as you try to have this thought (“I am now certain I am thinking, ‘The cat is on the mat’”), what you are actually thinking is that “I am now certain that I am certain that I am thinking, ‘The cat is on the mat.’” You’re engaged in another reflective task. You’re stepping back from the object that you *meant* to be wondering about (in the first case, “The cat is on the mat,” in the second case, “I am certain that the cat is on the mat”), you’re stepping back from that object and you’re engaging in a new thought (“I am certain—now—that I am certain that I am thinking, “The cat is on the mat’”). Well, I claim that certainty is going to remain elusive, because in the Cartesian sense, pure intellection lags a step behind its object, you can’t focus on the object that you need to focus on in order to figure out whether you’re certain. Hmm. What’s going on here?

Here’s another way of putting it: there’s a gap between two acts of thought. One is the thought that you’re having, that’s one act of thought, and another is the reflection vis a vis your certainty with respect to that first act of thought. And if there are two acts of thought—if there are two moments of pure intellection—then the possibility of error always looms. The evil genius could step in between the two acts of thought. Two acts of thought spatially, two acts of thought temporally, an evil genius could step in between them to deceive you. So you might well be wrong about what you’re thinking, because every time you focus on what you’re thinking, your focusing counts as a new thought that’s different from the thought that you hope to be

wondering about. Hmm. A gap between two acts of thought. Well, ok. Maybe, maybe I can persuade you of this. But some of you, including my colleague Will Dudley, some of you are going to respond as follows: what about being certain that you have thoughts? Just the *fact* of your thinking, even if you can't be certain about which one you are having. How about the claim, "I'm certain that I am now thinking, though I don't know which thought I'm having." The weird thing about this, Will, is it's puzzling how you would know that you're having a thought if you didn't know which one you were having. I mean, after all, if you can't tell which thought you're having, what persuades you that it's a thought rather than a bit of after-lunch biliousness, or something else quite different happening to you, if you can't tell which thought it is? So, when you face the challenge of understanding the act of thought, and trying to understand whether you can be certain of the act of thought even if you can't be certain of the content of it, I claim that in the absence of knowing which thought you're having, there's little reason to think that you're having any thought at all.

Now we're going to return to this, because there might well be a marker for thought, there might well be something about the very act of thinking that we can hold onto beyond the contents of a particular thought. That might well be. And I think Descartes is onto that possibility, and it's the possibility that I want to talk about next. So, we haven't—we're not done with this particular proposal. But, as first pass, I'm going to complain that if you don't know what thought you have, then it's puzzling, your confidence that you're having one.

So, just returning to our starting point, Descartes claims that he's certain that he exists as often as he explicitly thinks it (we're treating thinking as explicit right here). So Descartes claims that he's certain that he exists as long as he's explicitly thinking. But, he can't be certain how often that is, so he cannot be certain that he exists, I claim. I'm trying to undermine our

confidence in our own thoughts and leverage that argument into an argument against Cartesian certainty. He claims I know I exist as long as I am explicitly thinking, but I claim he doesn't know how often he is explicitly thinking so he can't be certain that he exists. Hmm.

There's another sense of thinking, though, that Descartes might be exploiting. We've been talking about thinking as this articulated, conceptual, reflective capacity, but there's another sense of thought that Descartes has been working with in the Meditations—thinking as what we might call *consciousness*. And when we reflect on *that* possibility, on the possibility that Descartes is emphasizing thinking as consciousness, we get a different set of challenges or we face a different set of challenges in determining whether Descartes is right. So, let's focus on thinking as consciousness. What's consciousness again? Well here's more of Descartes on the matter, going to show that he's thinking of thinking in at least two ways: Descartes says, "It is certain that I see to see light, hear a noise, and feel heat. This cannot be false, and this is what in me is properly called conceiving, which is nothing else than thinking." So, Descartes has another sense of thinking operating in his philosophical view. Another sense of thinking—thinking as consciousness. And here are some examples for you: there's tasting chocolate—one of my favorite of the conscious experiences—there's tasting chocolate, there's seeing the color red, there's feeling envy, there's experiencing pain. Conscious experiences. Visceral. Part of your repertoire in engaging the world. Having sensation. So here's another sense of thinking—not the articulate, explicit, conceptual stuff that we were talking about earlier. Thinking as consciousness, as we might say the felt sensation of your present experiences.

So, can you be wrong about how your sensations feel to you? Not the words—of course you could be wrong to use the word *chocolate* when in fact you are confronted by something much more like *tomato* in your mouth, you could be wrong about the word *chocolate*, but we're

not talking about the words. We're talking about the sensations themselves—could you be wrong about those, or could an evil genius deceive you with respect to the way your sensations feel? All right. So we're asking the same question about this sense of thinking.

Well here's an argument for how you might be wrong about your conscious experience—wrong about your conscious experience. Anybody tasting chocolate right now? Any lucky folks in the audience tasting chocolate? Well if you need a—if you need to do a little experiment here, you could very gently pinch the edge of your hand (don't pinch too hard, just pinch the edge of your hand). And you've got some sensation there, some sensation right there, a kind of "on the way to pain," depending on how indulgent you are in the philosophical experiment. So, could you be wrong about that? Well here's an argument for you being wrong about the sensation that you're feeling.

Being conscious is (some people say) a glorious, wondrous, mysterious capacity that you have, there's *being conscious.* You're able to do it, I presume. I have trouble finding out in your cases, but I'm going to work with this idea that it's a capacity you have. And it's a capacity you have that's likely shared with many non-human animals. A capacity that I expect that my cat had, it's a capacity that I expect bonobos to have, it's a capacity that I expect extends in the animal kingdom, *being conscious*. What's my evidence for this? It's indirect, to be sure, just as it is in thinking that *you* have consciousness, indirect evidence. But it's a capacity that's likely shared with many non-human animals. On the other hand, you have the capacity to reflect, reflecting on your consciousness in a *different* (glorious, wondrous, mysterious—possibly) capacity that you have. And it's a capacity that's likely *not* shared by many non-human animals—reflecting on consciousness, thinking about themselves as having a conscious experience, let alone thinking of themselves as being *certain* that they have a conscious

experience. So while I think that my cat had a conscious experience, had many conscious experiences (but for the time that he was sleeping), while I think that my cat had many a conscious experience, I don't think that he was able to reflect on his conscious experience, I don't think that he was able to reflect on it with the kind of sophistication that would be required for him to determine that he was *certain* that he was conscious. So, reflecting on consciousness is different from having it. Huh. But, if consciousness and reflecting on consciousness are two different capacities, then an evil genius could enter the gap to mislead you—two different acts of thought, two different moments of cognition. Being conscious, reflecting on consciousness.

On my laptop this is red—doesn't it look red to you? Well, at least we're all using the same color words. So, on my laptop this is red. You're having a conscious experience of the redness, well, you're having a conscious experience of the kind of washed-out redness. That's one thing, the consciousness of the redness. Here's another: I am *certain* I am now seeing red. You have this reflective capacity to claim that you're certain you're now seeing red. But they're two different moments, they're two different acts of thought, two different capacities that you have. You possess the capacity to do both; my cat likely only possessed the capacity to do one. If they're two different capacities, if there's a gap between them, a spatial gap or a temporal gap, an evil genius could enter the breach and deceive you. It's *conceivable* that you're misled in the second act, it's conceivable that you're wrong that I am certain I am now seeing red. Why is it conceivable? Because that first act of cognition, that moment of consciousness, might not *be*—it's something different from the reflection itself.

So—I'm elaborating for you here, Will—what about being certain that you're conscious even if you can't be certain what you're conscious *of*? Ok, I can't be certain that I'm seeing red, I can't be certain that I'm tasting chocolate, but how about being certain that I'm conscious *at all*?

How about being certain of that? Yeah, but if you can't be certain what you're conscious of, how can you be certain that you're conscious? Here's the thing about the explicit thinking case: remember ten minutes ago, when we entertained the possibility that you could be sure that you're *thinking* even if you couldn't be sure what you were thinking about? Remember, we entertained that possibility. Maybe what was lurking in your philosophical imagination at that point was that there's a *mark* to thinking that's different from the content, that's different from the explicit thing that you're thinking about. So maybe what was in your philosophic imagination was, there is something phenomenological, or having to do with conscious experience, that was the mark of having a thought. And maybe you thought, ten minutes ago, that you could be certain of something that had that marker, something that possessed that property, even if you couldn't be certain of the specific thought that you were having. And that wasn't a crazy thing for you to think ten minutes ago, to think that thinking in the explicit sense had some consciousness associated with it. Maybe it *feels like* something to think, for all I know—I don't do very much of it, so I'm not sure, but at any rate—I entertain the possibility that it feels like something to think. But in this case, now talking about consciousness, once you lose certainty with respect to that which you're conscious of, there's no other marker. The buck stops, as it were, with consciousness—there's no other hint about the presence of consciousness other than consciousness *itself.*

So, if you can't be certain of what you're conscious of, then how can you be certain that you're conscious? What would the other marker *be*? It's not consciousness anymore, and it's not explicit thinking because you can't be certain of that, and therefore I conclude that you can't be certain of explicit thought or of conscious thought. There's no other marker for the activity of consciousness than consciousness *of something*. And so certainty is elusive, because conceptual

reflection lags one step behind consciousness—the conceptual reflection of you *wondering* whether you can be certain of your conscious experience. There's a gap between an act of thought and consciousness, and in the gap lies the possibility of error. So, completing the second part of the argument, Descartes claims that he's certain that he exists as often as he—now *consciously*—thinks, but he can't be certain how often that is, so he cannot be certain that he exists. You can't be certain of your explicit thoughts, you can't be certain of your conscious states. The Cartesian argument fails, I claim. The best possibility for certainty are non-starters, I claim.

Here's the most reasonable question in philosophy: indeed, "So what?"

I got interested in this question—this question of certainty—because what I really wanted to think about was consciousness itself, what I really wanted to think about was understanding consciousness within a particular framework for understanding the world. For me, that framework is a sort of naturalistic or scientific framework. For me, philosophy is best done and best understood when it's a kind of preparation for empirical work. For me, philosophy is best conceived of when it's conceived of as continuous with what goes on in the sciences. So I got into the question of certainty because I was interested in the question of consciousness, and I was interested in how consciousness fits into the possibility of conducting a naturalistic exploration of our world, an exploration having to do with empirical approaches. That's how I got interested in this. And so for me, the answer to this question is as follows. Huxley (contemporary of Darwin, defender of Darwin) says the following in the *Lessons on Elementary Psychology*: "How is it that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of a djinn when Aladdin rubbed his lamp." How does consciousness fit into our naturalistic conception of the world?

Some of my tutorial students are here—I'm teaching a tutorial, working with students in a tutorial on consciousness—and I'm actually rather embarrassed to be admitting my views so early in the semester. (Luckily my views are so mad that they'll all believe that I'm kidding.) But maybe some of those students in the consciousness tutorial were attracted by the course description—which, I don't know what it says, I wrote it in some fit of enthusiasm—the course description, the course description says something like "consciousness looms as the greatest mystery confronting humankind since the beginning of time" . . . stuff like that, is what the course description says. And maybe they were intrigued by the course description because it seemed to them that consciousness really *is* a mystery to be contended with. Maybe it seemed to them that David Chalmers (the famous researcher on consciousness, an old friend from the University of Arizona), maybe it seemed to them that Chalmers is right when he says that consciousness remains as the last great terrain to be explored intellectually. Maybe it seems to them that Chalmers is right, just as it seemed to Huxley. I don't think so anymore.

Here's the alleged hard problem of consciousness, and here's my answer to the "So what?" question of the first half of the lecture, the first half of the lecture was to undermine some of our certainty with respect to thinking, both in the explicit sense and the sense of consciousness. So here's my answer to the "So what?" problem, and it has to do with consciousness. Here's the hard problem: it's the problem of experience, or the "something it is like to be" problem. It's . . . *something* . . . to be you, there's some way it is like to be you. I don't know what way that is—I'm sure it's much better than the way it is to be me. So, so there's some way that it's like to be you—you have sensations, you have experiences, your waking life constitutes something essential about you. And that doesn't seem to fit so easily into our scientific inquiry. It's a hard problem, and it's a hard problem why? Because it's alleged to have,

this “what it’s like to be you” problem or this “what it’s like to be you” phenomenon, is alleged to have scientifically intractable properties. Scientifically intractable properties. To start with, it’s alleged that consciousness—your consciousness—is epistemically unique.

Descartes again, from the Sixth Meditation (this is the paragraph, the ninth paragraph, where Descartes proves with certainty that the mind is separate from the body . . . . Hmm. Maybe. Well, it could be). So, here it is, the Sixth Meditation, Descartes writes, “Simply from knowing with certainty that I exist, and that meantime I do not observe any other thing as evidently pertaining to my nature, or to my essence, except that I am a thinking thing, I rightly conclude that my essence consists in this alone, that I *am* a thinking thing. And although possibly I have a body with which I am very closely conjoined, yet since on the one hand I have a clear and distinct idea of myself, insofar as I am only a thinking unextended thing, and on the other hand a distinct idea of the body, insofar as it is only an extended unthinking thing, it is certain that I am truly distinct from my body, and can exist without it.” Hmm. It is certain that I am truly distinct from my body. How does Descartes come to this ambitious certainty? How does he come to this dramatic conclusion? (Well, it’s like four sentences, and Descartes has shown with certainty that he is separate from his body.) How does he come to this conclusion? He comes to this conclusion on the basis of the alleged epistemic uniqueness of consciousness. He can be certain that he is conscious, he can be absolutely certain that he exists as a thinking thing. He has no such certainty with respect to his body—he might be quite wrong about his body, he might be deceived by an evil genius, he might be in the Matrix and be Keanu Reeves instead. You might be quite wrong (in fact, I often hope I’m quite wrong about my body). You can be quite wrong about body. You *can’t* be wrong about the mind, according to Descartes.

So, the mind—your knowledge of your own consciousness—is epistemically unique, claims Descartes. And that's part of the hard problem of consciousness: how is it that you can have this epistemically unique relationship to consciousness that you can't have to your body. Part of the hard problem of consciousness. But, uh, the thing is, we went over that one. Descartes' wrong. So, consciousness doesn't have this property that he alleges it to have. Consciousness isn't epistemically unique in the way that Descartes needs it to be in the Sixth Meditation.

So what's the hard problem of consciousness? What's so scientifically intractable about consciousness? Well maybe it's that consciousness is private: nobody can know what you are experiencing or thinking than you (a good thing, sometimes). Consciousness is essentially private. And this property of essential privacy, this property looks like it's going to render consciousness immune to scientific investigation, because scientific investigation—among many other things—scientific investigation requires that its object be in some sense public, studyable by many, intelligible by many, renderable by a community. Seems to be required in order to do science. But consciousness appears to be utterly and essentially private. And that's what makes consciousness such a hard problem, uh, a kind of make-work for philosophers. (Lucky thing). But, I don't think so. There's a gap between your reflection and the object—the object being whatever consciousness is. There's a gap between your reflection and the object, that was established in establishing that you can't be certain that you're conscious, there's a gap between your reflection of the thing and the thing that you're reflecting on. And that gap is not so different from the gap between someone else and the object, not so different from the gap that might exist between some other researcher and the object.

Descartes needs your reflection on consciousness to be related to your consciousness differently in kind than everybody else's relationship to your consciousness. He needs that difference to be different *in kind.* That's how he drives his metaphysical thesis, that's how he concludes mind-body dualism. If I'm right, the difference is—if there's any at all—a difference in *degree.* There's a gap between your reflection and the target of your reflection (in this case, your consciousness, whatever it is—I don't know, do you?), there's a gap between reflection and consciousness. So too there's a gap between *my* reflection and *your* consciousness. There's a gap. Fine, there's a gap—we need to figure out a way to minimize that gap. That's what science is—a systematic and glorious way of articulating concepts such that they render the object intelligible, such that they explain the object, such that they situate the object within a coherent and rich conception of the world. That's what science is. And the fact that there's a gap between the scientist and your consciousness is *no different* from the fact that there's a gap between your reflection and your consciousness. There are gaps, there are gaps, there are gaps.

So, what's the hard problem of consciousness again? Consciousness isn't epistemically unique, it isn't private in the sense that Descartes needs it to be private (it isn't private in that sense—there's a gap between your reflection and the target just as there's a gap between my reflection and you, your consciousness). So what's the hard problem? Here it is: it's that consciousness is perspectival. Consciousness requires that you be you in the experience that you're having. It has to do with point of view, it has to do with some angle on the world that you have. Consciousness is perspectival. And, the thing about science is, it's alleged to not be perspectival—not in the same way, anyway. The thing about science is that science doesn't require that you have the experience of lightning in order to understand what lightning is, science doesn't require that you have the experience of touching or tasting water in order to understand

what water is. You can understand those things perfectly well, without—in the form of science—without encountering them. In the famous thought experiment involving the fictional neuroscientist, Mary, we're to imagine Mary knowing *everything there is to know* about color vision. We're to imagine Mary knowing everything there is to know about the neurophysiology of human beings. But here's the thing, in this thought experiment: imagine that Mary got this knowledge in a black and white room that she grew up in. So here's Mary, knowing everything there is to know scientifically (this is fictional, we'd never actually do this to Mary), so imagine Mary knowing everything there is to know about the human nervous system—with an emphasis and specialization on the color vision system—imagine that she knows everything there is to know, but she grew up in a back and white room. The claim is that when we take her out of the room, and show her a bright red tomato, the claim is that she knows something new, she has a new *perspective* on this bit of knowledge. Consciousness is perspectival: it requires that you have the experience. And *that's* the hard problem of consciousness (allegedly).

But, I claim that understanding is conceptual and reflective. Your understanding of your own consciousness is conceptual and reflective. And so, it's not perspectival in the sense that the friend of consciousness needs, because the friend of consciousness in the Mary argument is playing a trick on you, I say. She's playing a trick on you, because what she was trying to get you to do was focus on the redness of the tomato, she's trying to get you to focus on the redness of the tomato, without remembering that Mary has some conceptualization, some reflection, about the redness of the tomato. That's what understanding *is*. So, the friend of consciousness was trying to get you to ignore the conceptual reflective part in favor of something else, something else we know not what (I don't know, nor do you, because that's the target of our project here, trying to understand consciousness). She's trying to get you to reflect on that, or

concentrate on that, without remembering that understanding is conceptual and reflective. And, understanding is conceptual and reflective in your hands, in Mary's hands, and in our hands when we're engaged in trying to understand what consciousness is. So if there's a perspectival component to understanding consciousness, then that perspectival component is one that troubles *both* the first person *and* the third person perspective on consciousness, because we're talking about understanding here, we're talking about having—we're deploying this reflective capacity. So, perspectival capacities having to do with consciousness are not the intractable properties that friends of consciousness claim it is.

So I say that the properties allegedly had by consciousness are the result of a fallible inference based on particular evidence. You think consciousness is some way—I don't know what way you think consciousness is, we'll talk about it in five minutes. I don't know what you think of consciousness, but you think that consciousness has some properties, and you think it has those properties because, you know, you're smart, and you've reflected your whole life on your consciousness, and you think it has some properties that need to be made sense of. Fair enough. That's what scientists are always faced with: a phenomenon that has some properties that need to be made sense of. Those properties—in the case of science—those properties are ones that are arrived at through fallible inference. Scientists can be wrong. So too might you be wrong with respect to the properties had by your consciousness. You might be wrong. How might you be wrong? You got the story incorrect. You were persuaded by culture, you were persuaded by reading Descartes in philosophy 102 (leave while you can, all right—I'm sure drop-add hasn't ended yet), you were persuaded by what your friends said about consciousness, you were persuaded by folk stories having to do with consciousness, you'll be persuaded by reading David Chalmers' book in the philosophy tutorial, you were persuaded by some *New York*

Times piece by John Searle regarding the mysteriousness of consciousness, you were persuaded by something or other about the properties of consciousness. But, you're fallible. You might be wrong that it has those properties. And the properties that are alleged to be intractable, the properties that make consciousness allegedly the hard problem are ones that I think we're in fact wrong about.

I don't know what consciousness is. But I think we're wrong to think that consciousness has properties that make it immune to scientific investigation. And what I'm hoping is that as scientists, or as proto-scientists (which is how I understand philosophy), I'm hoping that as scientists we can move forward with making sense of the properties of consciousness in such a way that we can come to an understanding of it within the same framework that has been so wildly and massively successful in the last 400 years of human history.

So I've been urging philosophy as continuous with science, and that continuity is what motivated me to start thinking about Cartesian certainty. I was worried: if Descartes' right, if Descartes' right to think that we have a different epistemological relationship with our mind than we have with anything else in the world, then he's on to something when he claims that the mind is different than the body, that the mind is not a subject matter for science but rather for speculative philosophy. If he's right about that epistemic difference, then he's on to something. I think he's not right about that epistemic difference. I want to treat philosophy as continuous with science, as preparation for empirical inquiry. I want to sign on to a view of philosophy defended by (among many others) the late W. V. O. Quine, when Quine says in *Epistemology Naturalized* that epistemology—the study of knowledge—simply falls into place as a chapter of psychology, and hence of natural science. Epistemology, or something like it, studies a natural phenomenon: a physical human subject.

"Knowing One's Mind" didn't originate in my mind by myself—philosophy, like science, for me, emerges best in collaboration and conversation. Here are some folks that stand out as being folks that I've learned from over the last couple of years in thinking about these topics (some of you are in the audience, thanks a lot, and thanks all of you for coming).*

* Many thanks to my research assistant Rachel Schneebaum for transcribing the lecture and for conversation on these topics.