Knowing Our Minds

Why some philosophers say we can't

ALEX BYRNE

"Know thyself" was inscribed on the ancient Greek temple at Delphi and is quoted approvingly by Socrates in the *First Alcibiades*. What is it to know oneself—to have self-knowledge? It is, at least, to know facts about oneself. And, in general, knowing facts about oneself is not hard. For example: I know that I live in Cambridge, that I am now sitting at a desk, that I have a twinge of pain in my right leg, that I prefer red to white. But that is not the sort of knowledge the Delphic motto enjoins me to cultivate. According to Socrates, to obtain the right sort of knowledge about oneself one must "look at the soul; and especially at that part of the soul in which . . . virtue resides," which seems a little harder than finding out one's address. This important kind of self-knowledge concerns who I "really am" as a person, or my "essential nature," or something along equally lofty lines. Perhaps such self-knowledge is unattainable—if Sartre is right, there is no essential nature to know.

The phrase "self-knowledge" tends to evoke the Delphic kind of knowledge of oneself, and it is freely sprinkled around the metaphysical and self-help sections of bookstores. Contemporary philosophers understand the phrase differently: for them, "self-knowledge" is simply knowledge of one's own mental states. This sort of knowledge—our topic—is commonplace: knowledge of what one is thinking, of what one wants or intends, of one's sensations. Intoxicating Sartrean doctrines such as "existence precedes essence" can be discussed when the calvados is finished: as we shall see, the humdrum kind of self-knowledge is puzzling enough.

How do I know that you have a pain in your leg? Perhaps you tell me or I see you hopping around or grimacing while holding your leg. In short, I know that you have a pain in your leg by observing your behavior (including your verbal behavior). Of course, sometimes this method doesn't work because you're faking—but usually it does. How do I know that *I* have a pain in my leg? In the typical case, not by asking my doctor or seeing myself in the mirror hopping around. It is not immediately clear how best to characterize the method I

normally use to find out whether I am in pain, but it is clear that whatever that method is, it is quite unlike the way I have of knowing that *you* are in pain. And similarly for other mental states. I know that *you* believe that the pub is open because I see you striding purposefully toward it; I don't know that *I* believe that the pub is open by catching a glimpse of myself in a store window, heading for the pub. Usually I do not need to observe myself to find out what I believe. A person, then, has a special way of finding out about her mental states that is quite different from the way she finds out about others' mental states. Let us mark this fact by saying that we have *peculiar access* to our mental states. Contrast this with finding out one's own weight or underwear color: here all the methods of discovery—using scales, undressing, and so on—can also be employed to find out these facts about other people.

Consider another question: can you think you're in pain but be wrong? That is, might you falsely believe that you're in pain, as you might falsely believe that you're wearing hot-pink boxer shorts? Some philosophers think not: if you believe you're in pain, you must be in pain, they say. But even if errors about pain are possible, they seem in general rarer in the first-person case: I can easily falsely believe that *you* are in pain, but typically when I believe that *I* am in pain, I really am in pain. And similarly for other mental states. I can easily be wrong about whether you believe that the pub is open. More seriously, I can easily be wrong about whether the pub is in fact open. But typically when I believe that I *believe* that the pub is open, I really do believe that the pub is open. That is, our beliefs about our mental lives are less prone to error than our beliefs about others' mental lives and our beliefs about our external surroundings. Put more or less equivalently: knowledge of one's own mind is easier to obtain than knowledge of most other subject matters. Let us mark this fact by saying that we have *privileged access* to our mental states.

Although it is fairly uncontroversial to say that we have privileged access to our mental states in some form or another, its extent can be exaggerated. With the possible exception of sensations such as pain, errors about one's mental states are not at all uncommon. Do I feel resentment, or is it just irritation? Do I really believe that *The Da Vinci Code* is a shoddy potboiler, or do I just give that impression to avoid seeming lowbrow? Sometimes others are much better placed to find out what your mental condition is, and their discoveries can cost you hundreds of dollars an hour. In the early chapters of George Sand's *Marianne*, it is apparent to the reader that Pierre loves the eponymous heroine, but

Pierre himself believes the opposite. And—as summarized in Timothy Wilson's recent book *Strangers to Ourselves*—social psychology has uncovered our remarkable tendency to confabulate about our mental lives.

Peculiar access—a special first-person method of knowing our own minds—and privileged access—easier knowledge of our own minds—are quite different features of self-knowledge: the first could be present without the second, and vice versa. The British philosopher Gilbert Ryle agreed that we have privileged access to our mental states, but denied that we have peculiar access. Ryle thought that we find out about our mental states in the same way that we find out about the mental states of others—by observing behavior. Thus, in Ryle's view, we lack peculiar access. But we do have privileged access, according to Ryle, simply because we have much more evidence of our own behavior than that of others; we are our own constant companions, after all. "The turns taken by a man's conversation," he writes in *The Concept of Mind*, "do not startle or perplex his wife as much as they had surprised and puzzled his fiancée, nor do close colleagues have to explain themselves to each other as much as they have to explain themselves to their pupils." Privileged access, according to Ryle, is no more problematic than the fact that one's spouse and close friends are less likely to err about one's mental life than one's casual acquaintances—the explanation of both is basically the same. You spend so much more time with yourself than anyone else does—how could you fail to be the best authority?

Conversely, one could consistently maintain that we have peculiar access while denying that we have privileged access. To see that this is a possibility, note that we have a kind of peculiar access to our own bodies. I often know that my legs are crossed, not by looking, but by exercising my faculty of proprioception—a special way of perceiving the disposition of one's body. I cannot use proprioception to find out whether *your* legs are crossed. And similarly for you: each of us has a first-person method of discovering his or her posture. But this method is not immune from error. Proprioception can lead one astray, and sometimes looking is a better guide. Similarly, perhaps the peculiar access we have to our own mental states is not especially reliable.

Although this does show that there is no *logical* relationship between privileged and peculiar access, the orthodox view—against philosophers like Ryle—is that self-knowledge

does indeed have these two features. If this is right, the chief philosophical problem of self-knowledge is to explain how we have privileged and peculiar access to our own mental states.

But why is this problem particularly hard? As we shall shortly see, both features of self-knowledge are as mysterious as they are familiar. Let us begin with a much-discussed difficulty for peculiar access—sometimes called *the puzzle of armchair knowledge*. In order to do that, a small quantity of philosophical machinery needs to be wheeled in and explained.

In the first of his *Meditations*, Descartes wonders whether there are grounds for doubting his beliefs, for instance the belief that he is sitting before his fire. To that end, he considers the possibility that a "malicious demon" has contrived to produce in him the perceptual experience of sitting by a fire, even though there is no fire anywhere, and likewise for all of Descartes's perceptual experiences. If there is such a demon, "the sky, the air, the earth, colors, shapes, sounds, and all external things are merely the delusions of dreams." However, Descartes takes for granted that the demon hypothesis makes no difference to his mental life—in particular, to his thoughts and beliefs. Even if there is a malicious demon, and therefore no fire nor any "external thing," this is perfectly compatible, Descartes apparently thinks, with his *believing* that he is sitting before the fire. His thoughts and beliefs would remain the same whether he was the victim of an evil demon or not.

Descartes implicitly assumes that one's "inner" mental life is fundamentally independent of how things are in the "external world"—in one's surroundings or environment. And this can seem like sheer common sense, especially to the contemporary ear. After all, the brain is the seat of the mind, and the brain is entirely in the head. So shouldn't the mind be *literally* "inner"—that is, also entirely in the head?

This Cartesian doctrine of a boundary between the inner mind and the external world can be made more precise. Here is a way of doing this—not one congenial to Descartes, but that won't matter. Imagine a person who is a perfect replica of you from the skin in—a

molecule-for-molecule duplicate, exactly the same in every detail. Such a person would be internally just like you, and according to the Cartesian doctrine such a person would have to be mentally just like you—in particular, your replica would have the same thoughts and beliefs. Let us call this claim *internalism*.

Of course your body double need not be like you in every respect—only in internal respects. For example, suppose you are a mile away from the nearest pub, have some beer in your fridge, have a zero blood-alcohol level, and believe (among many other things) that wine has health benefits. A molecule-for-molecule replica of you need not be a mile away from the nearest pub, or even have a fridge, let alone one with beer in it. On the other hand, your double is guaranteed to have a zero blood-alcohol level, since the two of you are exactly alike with respect to the distribution and arrangement of molecules. If there is no alcohol in your blood, there won't be any in your double's. Now, what about the belief that wine has health benefits? Is having that belief an entirely internal matter, like one's blood-alcohol level? Yes, according to internalism: given that you have this belief, your double is guaranteed to have it too.

Until recently, internalism was tacit philosophical orthodoxy. But in the 1970s it came under severe attack, principally from two philosophers, Hilary Putnam and Tyler Burge. To cut a long story short, Putnam and Burge argued that the thoughts and beliefs of your double could—if we fill out the details of the story correctly—be different from yours. That is, what one thinks and believes doesn't just depend on what goes on *in* the head—it also depends on what goes on *outside* the head, in one's environment. (Some philosophers have subsequently suggested that this is also true for sensations, such as pain, but we can set this aside.)

To get the flavor of Putnam's and Burge's arguments, consider another of your beliefs—say, that Larry Summers is a controversial university president. Will any perfect replica of you also believe that Larry Summers is a controversial president? Imagine your double—Twin You, we'll call him or her—living on a distant planet, pretty much like Earth—Twin Earth, we'll call it. On Twin Earth there is a university superficially like Harvard, with a president who is as controversial as our own planet's Larry Summers, and who could pass for him in a police lineup. This otherworldly president is even called "Larry Summers" by the inhabitants of Twin Earth; to avoid confusion, we'll call him Twin Larry. (Note: It is

not important to imagine that Twin Larry is a perfect replica of Larry—there can be all sorts of differences, provided that you and Twin You remain perfect replicas.) Now, Larry—the president of Harvard—has always been Earthbound, never venturing beyond our solar system. And Twin Larry has never been near Earth, having lived all his life in a galaxy far, far away. They are therefore distinct albeit similar individuals, like my particular copy of the last issue of *Boston Review* (in my bathroom) and yours (on your coffee table).

Internalism, remember, is committed to saying that you and Twin You both believe that Larry Summers is a controversial president. But is that right? Does Twin You believe that *Larry Summers* is a controversial president? Of course not. Twin You has never visited Earth and has never heard of Larry Summers—Twin You has never met him, read about him, or glimpsed him on the street. Rather, Twin You has heard of *another* individual entirely, namely Twin Larry, who also happens to be called "Larry Summers." Since Twin You has never heard of Larry Summers, Twin You does not believe that Larry Summers is a controversial president; instead, he or she believes that Twin Larry is a controversial president. Here we have two people—you and Twin You—who are internally exactly alike, but who have different beliefs! Whether Twin You's belief is true does not depend at all on Larry's penchant for putting the cat among the pigeons but on Twin Larry's. In fact, with a little more elaboration we can tell the story so that your belief is true and Twin You's is false.

By means of more sophisticated and complicated versions of thought experiments like the one just described, Putnam and Burge convinced most philosophers that internalism is false. To adapt a famous quotation of Putnam's: cut the pie any way you like, thoughts just ain't in the head! And not just thoughts and beliefs about individuals, such as the belief that Larry is a controversial president. Internalism is false for practically every belief—including, for instance, the belief that wine has health benefits. In general, what one thinks about and believes is partly a matter of what goes on outside the head: *externalism* is true.

(Now that we've gotten the difficult case of belief out of the way, we can note that internalism is clearly false for some other mental states—loving, for example. Pierre loves Marianne—but one can only love Marianne if she is there to be loved, just as one can only kiss or kick Marianne if she is there to be kissed or kicked. If we imagine Pierre's perfect

replica on a Marianneless distant planet, he would not love *her*.)

This just skims the surface of Putnam's and Burge's case for externalism, and it ignores a labyrinthine collection of objections and replies. But it suffices to explain the puzzle of armchair knowledge, which was first sharply formulated by the philosopher Michael McKinsey.

Remember that this is supposed to be (primarily) a puzzle for one of the two features of self-knowledge: peculiar access. Whatever having peculiar access amounts to, it does not seem to involve observation of our environment or surroundings, as does our knowledge of others' minds. As Richard Moran puts it in his *Authority and Estrangement*, "A person can know of his belief or feeling without observing his behavior, or indeed without appealing to evidence of any kind at all." Merely by sitting in an armchair and considering the matter, one can come to know that one is thinking of Larry, or that one believes Larry to be a controversial president. But the Twin Earth thought experiments show that one can have beliefs or thoughts about Larry Summers only if one has met him, seen him, otherwise perceived him, or heard about him from other people—only if, let us say, one has *encountered* him. And whether one has encountered Larry cannot be settled by armchair-bound methods alone.

So here—at last—is the puzzle. Suppose that you are idly thinking of Larry Summers. Because of the peculiar access you have to your own mental states, you can know without past or present observation of your environment or behavior that you are thinking of Larry. Still in the armchair, you can also run through some philosophical thought experiments and learn that one can only think of Larry if one has encountered Larry. Putting those two pieces of knowledge together, you can conclude—again without ever rising from the armchair—that you have encountered Larry. But, as noted in the previous paragraph, surely you *cannot* come to know that you have encountered Larry by armchair reflection: ordinary empirical investigation is required.

In short, the overthrow of Cartesian internalism threatens to undermine the idea that we have peculiar access—in particular, "armchair" access—to our thoughts and beliefs.

Turn now to the second feature of self-knowledge—privileged access. In order to explain a problem for privileged access, we need to introduce a perennially popular suggestion about the mechanism of self-knowledge: a kind of "inner" perception.

We discover how things are in our environment by using our eyes, nose, and so on. We see the cat and thereby come to learn that it is on the mat, smell the coffee and thereby come to learn that it is brewed. Why not say that we discover how things are in our minds by similar means—by employing a perceptual faculty specialized for the detection of beliefs, intentions, sensations, and so forth, as vision is specialized for the detection of colors and shapes? After all, the word "introspection" comes from the Latin for *to look into*, and talk of "the mind's eye" sounds perfectly natural. (Recall that in the *First Alcibiades* Socrates speaks of "looking at" the soul.) And how else—apart from using our familiar senses to observe our behavior—could we find out about our mental states? Some sort of perception-like faculty—an "inner eye"—seems the only available option.

This inner-sense theory has prominent contemporary defenders. On the plus side, it offers an explanation of peculiar access (assuming that the puzzle of armchair knowledge can somehow be solved). We earlier noted that proprioception provides a way of knowing of the disposition of one's limbs that works only in the first-person case—one can't use proprioception to find out whether someone else has her arm raised. The inner-sense theory would explain *peculiar* access to one's mental states in the same style.

Unfortunately, the inner-sense theory leaves *privileged* access a complete mystery. As noted earlier, proprioception does not give one especially privileged access to the disposition of one's limbs. So, if there is an inner sense, why is it so much more reliable than our other senses? No one has ever satisfactorily answered this question.

That is the problem posed by inner sense for privileged access—inner sense is just another sense, so why is it significantly more privileged in its access to the mind than sight is in its access to external objects? Admittedly, this might be taken as a reason to reject the inner-sense theory rather than as a strike against our having privileged access. And, in fact, many philosophers have convinced themselves on other grounds that the inner-sense theory is a thoroughly misguided idea. One prominent objection is that self-knowledge is not as isolated from the rest of cognition as it would be if the inner-sense theory were true.

The loss of a perceptual faculty is a serious impediment, but it usually leaves one's rationality and intelligence intact. The deaf and blind are just as good as solving problems as the rest of us—they simply lack a conduit dedicated to delivering certain information about their surroundings. So, if there is an inner sense, one should be able to lose it while remaining a normal rational person. But there is reason to think that this is *not* possible. The philosopher Sydney Shoemaker has argued that a person lacking self-knowledge would be severely rationally impaired, in which case inner perception cannot be the source of self-knowledge.

Unfortunately, to reject the inner-sense theory is to throw the mezcal out with the agave worm. Arguably, the inner-sense theory is the *only* way of explaining how one can have self-knowledge "from the armchair"—without observation of one's past or present behavior. And much putative self-knowledge is from the armchair—knowledge of one's present thoughts or sensations, for instance. So, if the inner-sense theory is false, it follows that we usually do *not* know what we are thinking or feeling. But we do.

Let us sit back in the armchair and pause to take stock. The initially compelling claims about self-knowledge—that we have peculiar and privileged access to our mental states—collide with externalism and inner sense, respectively. That should prompt an examination of our starting point: maybe we were on the wrong track when we assumed that our relationship to our mental states is typically a matter of knowledge. It certainly appears as if "I have a pain in my foot" is used to report a discovery, something one has come to know, much as "I have a hole in my sock" is. But perhaps appearances are deceiving. Perhaps when you say "I have a pain in my foot" or "I am thinking of Larry Summers," you are not stating a fact about your own mental life, which you know through some strange method and with unimpeachable security.

What is the alternative? A suggestion can be found in Wittgenstein's *Philosophical Investigations* (although it probably would not be endorsed by Wittgenstein himself). Words like "pain," Wittgenstein writes, "are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations, and, later, sentences. They teach

the child new pain-behaviour." According to this view, to say "I have a pain in my foot" is to do something akin to holding one's foot and moaning. A moan is not true or false—the moaner is not reporting or describing anything. Moaning is a natural expression of pain, just as blushing is a natural expression of embarrassment.

Here we have a sketch of a theory—*expressivism*—that promises to dissolve the problem of self-knowledge. According to this view, we are misled by grammar into thinking that "I am in pain," "I am thinking of Larry Summers," and the like are typically used to report facts one has come to know about one's psychological life. Actually, "I am in pain" is not typically used to report or state anything, and is in this respect like the sentences "Pass the salt" and "Is there any more beer?" According to the expressivist, a more perspicuous paraphrase of "I am in pain," as that sentence is typically used, would be "Ow!" If you drop an anvil on your foot and shout "Ow!," you are not asserting that you are in a certain psychological condition—someone cannot sensibly reply by saying "That's true" or "I disagree." The distinctive kind of self-knowledge that leads to intractable perplexities turns out not to exist.

That sounds pretty radical—in fact, too radical to be remotely plausible. Here is one difficulty. When I say "You are in pain," I am describing your mental condition. There is no need for the expressivist to deny this platitude, because the problem is not with the idea of mental facts as such but rather with the idea of a special first-person knowledge of such facts. So "You are in pain" describes your mental condition—truly, let us suppose, in which case I am reporting a fact about you. But now notice that you can repeat back to me what I just said by changing the pronoun appropriately: "Quite right, I am in pain." In other words, by saying "I am in pain," you are stating what I just stated when I said "You are in pain." But this means that expressivism is incorrect, because according to it "I am in pain" is not in the fact-stating business at all.

Because of problems like these, a simple-minded expressivism about self-knowledge is not viable. But recently, various sophisticated descendants of the original view have been formulated. Typical utterances of "I am in pain" and "I am thinking of Larry Summers" (sometimes called "avowals") are made without empirical investigation and are rarely questioned or objected to. The simple-minded expressivist thinks that these distinctive features of avowals can be explained without supposing that we have privileged and

peculiar access to our mental states. In the case of "I am in pain," the explanation is that this sentence functions much like "Ow!"—which, as we have seen, is just *too* simple-minded. But there might be other and better explanations, and that is exactly what the thoroughly modern neo-expressivist holds. Dorit Bar-On's *Speaking My Mind* is the most developed version of a neo-expressivist account, which purports to offer "an account of avowals' special security that does not take it to be a matter of some specially secure epistemic basis or method." Bar-On does not deny, though, that an utterance of "I am in pain" or "I am thinking about Larry Summers" makes a straightforward factual claim about one's mental condition.

Interestingly enough, something similar happened to the analogous "expressivist" view in ethics, according to which "Stealing is wrong" is not used to report some ethical fact but rather expresses the speaker's disapproval of stealing. Modern neo-expressivist theories of ethics (as developed, for example, by the British philosopher Simon Blackburn) bear very little relation to the simple sort expounded early last century by A.J. Ayer and others, but they still find some insight in it.

It is sometimes said that philosophy deals with enduring questions that have preoccupied thinkers since before the time of Socrates: in philosophy, *plus ça change, plus c'est la même chose*. Science, on the other hand, moves relentlessly on, burning its bridges and consigning its early pioneers to the garbage heap of history. There is some truth to this—Aristotle's astronomical speculations are dated in a way his *Nicomachean Ethics* is not. But the central problems of philosophy change, too. In the tradition inherited from Descartes, the great epistemological questions concern our knowledge of the external world—how or whether one knows one is sitting before Larry Summers, and so forth. In this tradition, knowledge of our own minds—that one is *thinking* of Larry—is considered relatively unproblematic. Arguably, it is one of the singular achievements of contemporary philosophy to have shown that this has things exactly back to front.

More In Science & Nature

W H M U B C T W Minding
(S C tl ir ir E L Matter
W W O S H o o D MATTHEW

More In Philosophy & Religion