

Handout 9: Externalism and Self-Knowledge

Externalism and privileged access

So far, we have seen some arguments in favor of externalism -- the view that what mental states one is in is not in general an intrinsic matter. According to the externalist, two people can be internally exactly alike, and yet differ mentally. For example, Hilary here on Earth might believe that Cambridge (our fair city) is pretty, while his molecule-for-molecule replica Twin Hilary on Twin Earth, living in Twin Cambridge, does not share this belief. Instead, Twin Hilary believes that Twin Cambridge (a city many light years from our own) is pretty. We also discussed other examples with the same moral due to Putnam and Burge, and the "active externalism" defended by Clark and Chalmers.

The question now before us is this. Is externalism compatible with the (apparent) fact that each of us has a special way of knowing about our own mental states? According to McKinsey, it isn't.

In philosophical jargon, "self-knowledge" is knowledge of *one's own mental states*. For example, I know that I believe that I live in Cambridge; I know, in other words, that I am in the mental state of believing that I live in Cambridge, and so this is a piece of self-knowledge. I also know that I live in Cambridge; this is also a piece of knowledge about myself, but (in philosophical jargon) does not count as *self*-knowledge. It is quite natural to suppose that self-knowledge is radically different from our knowledge of the mental states of other people, at least in typical cases. I know that you are in pain, or that you want some lunch, or that you are wondering where to go for lunch, by observing your behavior (including verbal behavior). But I don't (typically) know that I am in pain, or that I want some lunch, or that I am wondering where to go for lunch, by observing my behavior. Further, I can easily be wrong about your mental states -- perhaps you're faking believing in pain, or perhaps you are heading off to the student center because you really want to buy a newspaper from LaVerde's Market, instead of a delicious offering from the food court. But I can't easily be wrong about my mental states -- if I believe that I am in pain, or that I want some lunch, then I'm usually right. (Perhaps I *can't possibly* falsely believe that I'm in pain, but for present purposes we can leave this issue aside; as discussed in class, we certainly can be wrong about some of our mental states.)

McKinsey puts the above points by saying that we have *privileged access* to our own mental states: "each of us can know the existence and content of his own mental states in a privileged way that is available to no one else". This claim has two distinct components: (a) that a way we have of knowing about our own mental states is *privileged* and (b) that this way is available to no one else:

(a) **privileged way**: it is (at the least) harder to be wrong about one's own mental states than about one's environment or the mental states of others.

(b) **available to no one else**: merely by sitting in an armchair, one can know, just by considering the matter, that one is thinking that water is wet, for example; one cannot know that someone else is thinking that water is wet by the same method. Although this is not explicit in McKinsey's paper, the conflict with externalism arises because of (b), not (a).

McKinsey puts the claim that one can know that one is thinking that water is wet merely by sitting in an armchair and considering the matter, as follows: one can know *a priori* that one is thinking that water is wet. It is, however, not entirely clear whether this use of the "a priori"

terminology is appropriate (this issue was mentioned in class). For safety's sake, let us avoid this jargon when stating (b), and put it instead by saying that one can know "from the armchair" that one is thinking that water is wet.

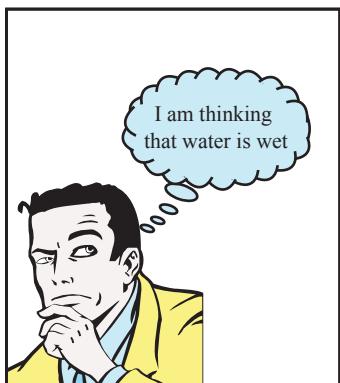


Image by MIT OpenCourseWare.

McKinsey's argument

McKinsey starts with the following argument template:

1. Oscar knows from the armchair that he is thinking that water is wet
2. The proposition that Oscar is thinking that water is wet implies E (according to externalism)
hence:
3. E can be known from the armchair

As McKinsey notes, for this argument to be valid, 'implies' in 2 needs to be interpreted as *conceptually implies*, which may be explained as follows: P conceptually implies Q iff it is knowable a priori that it is not the case that P is true and Q is false. That is, it is knowable a priori that if P is true, Q is true. Given this reading of 'implies', if 2 is true then Oscar can know a priori that if he is thinking that water is wet then E is true. Whatever "from the armchair" comes to, exactly, this much is right: knowing something a priori is one way (maybe not the only way) of knowing it from the armchair. So, since Oscar can know a priori that if he is thinking that water is wet then E is true, he can know this from the armchair. And since according to 1, he knows that he is thinking that water is wet from the armchair, he can know E from the armchair.

McKinsey then invites us to agree that there is choice of E such that: (i) 2 (so interpreted) is true if externalism is, and (ii) the conclusion 3 (so interpreted) is false. And if there is such an E, and assuming that 1 is true, by "reductio ad absurdum" 2 is false. And if 2 is false then externalism is false.

Brueckner on E

But what is E? McKinsey never says. Brueckner has some suggestions. The problem with all of them, he thinks, is that either the externalist will deny that the proposition that Oscar is thinking that water is wet implies E, or else it is not clear why E can't be known from the armchair.

Suggestion 1

E is the proposition that Oscar inhabits an environment containing H₂O and not XYZ.

Certainly one cannot know E from the armchair. And it is true that Putnam's Twin Earth thought experiment shows that one can only think about water if there is H₂O in one's environment, at least according the externalist. But does the thought experiment show this a priori? (For if not, then the externalist will deny that the proposition that Oscar is thinking that water is wet conceptually implies E.) No, it doesn't. In order to draw that conclusion from the thought experiment, we need the hard-won empirical assumption that the stuff that comes out of taps, falls as rain, etc., is H₂O. (N.B. it's vital to distinguish the proposition that water is such-and-such

from the proposition that H_2O is such-and-such. It seems that they can't be the same proposition, because it is possible to believe one but not the other: chemical ignoramuses believe that water is such-and-such but do not believe that H_2O is such-and-such.)

Suggestion 2

E is the proposition that water exists.

Certainly one cannot know E from the armchair. But will the externalist accept that Putnam's Twin Earth thought experiment shows (a priori or not) that one can only think about water if water exists?

She doesn't have to. Suppose hydrogen and oxygen exist, but hydrogen hydroxide doesn't, and that scientists "theorize that H_2O exists" (p. 641). Suppose that they introduce a term, *swater*, for this chemical compound, and use it on *Nova* broadcasts, in *Scientific American* articles, etc. Oscar reads these articles and learns the new word (perhaps without remembering the chemical composition of *swater*). Oscar might say, "I wonder whether *swater* is wet". Wouldn't he be wondering (in a waterless world) whether water is wet? More to the point, there is nothing preventing the externalist from agreeing.

Image removed due to copyright restrictions.

Picture of Element 114.

Suggestion 3

E is the proposition that either water exists or some in Oscar's speech community theorize that H_2O exists.

Again, one cannot know E from the armchair. The externalist will (we may suppose) accept that the truth of E is a necessary condition for Oscar to think that water is wet. But she will not accept that E is conceptually implied by the proposition that Oscar is thinking that water is wet (see the reply to Suggestion 1).

Suggestion 4

E is the proposition that there exist some physical entities distinct from Oscar.

Brueckner's response to this suggestion (the final one he considers) is simple: are we so sure that this E is not knowable from the armchair?

Suggestion 5

Brueckner seems to have overlooked another candidate for E, namely the proposition that either water exists or some in Oscar's speech community theorize that water exists.

This E is (arguably) conceptually implied by the proposition that Oscar is thinking that water is wet. (N.B. "theorize that *water* exists", not "theorize that H_2O exists"; this makes all the difference. This point is not easy to grasp, so you will need to work at understanding it.) Yet surely this E is not knowable from the armchair.

We are left with a puzzle, which (in this class, at any rate) will remain unresolved.

MIT OpenCourseWare
<http://ocw.mit.edu>

24.09 Minds and Machines

Fall 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.