Paradoxes of externalism and self-knowledge

January 20, 2004

1 The idea behind the arguments

Since externalism became prominent in the philosophy of mind, philosophers have worried that it might threaten the view that we have privileged access to the contents of our own minds.

There are several ways to see that there is a kind of intuitive tension between externalism and the doctrine of privileged self-knowledge.

1. One of the claims implicit in the view that we have privileged access to our own thoughts is that we can have knowledge of these thoughts without doing special investigation of our environment. But externalism claims that which thoughts we have our partly determined by facts about our environment; so how could we know about our thoughts without knowing the relevant environmental facts?

2. Imagine Oscar and his twin, Toscar, on twin-earth. There is a sense in which things *seem just the same to them*. In particular, their thoughts seem just the same to them. But they have different thoughts, if externalism is true. But, since their thoughts are indistinguishable to them, how could they tell this?

These intuitive challenges are both too vague to be convincing as they stand. But each has been turned into a more precise argument to challenge the orthodox view in the philosophy of mind that (i) some mental properties are externalist, and (ii) we have a kind of privileged self-knowledge of these properties. We will discuss these two arguments — (1) corresponds to what I call below the argument from illicit a priori knowledge, whereas (2) corresponds to the argument from slow switching.

You should keep in mind that, so far, we are working with a rather vague notion of 'privileged access.' It is worth asking what sort of privileged access each of the following arguments requires.

2 The argument from illicit a priori knowledge

(found in Boghossian, 'What the externalist can know a priori'; an earlier version can be found in McKinsey, 'Anti-Individualism and Privileged Access' (1991), in

Analysis. A good but complicated discussion of the paradox in the *Knowing Our Own Minds* volume is the McLaughlin & Tye contribution, 'Externalism, twin-earth, and self-knowledge.')

The first argument for the *incompatibilism* of externalism and self-knowledge has the form of a reductio: it says that *if* both externalism and the thesis of privileged access were true, *then* we could have a priori knowledge of certain contingent matters of empirical fact. But the idea that we could have such a priori knowledge is absurd; therefore one or both of externalism and the thesis that we have privileged access to our own mental states is false.

2.1 Boghossian's version of the argument

Boghossian states the following argument:

1. If I have the concept *water*, then water exists.
2. I have the concept *water*.

C. Water exists

Notes on the argument. 1. The distinction between meaning and reference. 2. What is meant by the expression 'the concept *water*.' 3. What is meant by talk of 'having a concept.'

2.2 The paradox generated by Boghossian's argument

It is important to be clear about why this argument is supposed to be interesting (and paradoxical). If you believe that externalism is true, then (Boghossian thinks) you must believe that (1) is true. And of course (2) is true; we have thoughts involving the concept we express by using the word 'water.' And (C) follows from (1) and (2); so the argument is sound. *But by itself there is nothing paradoxical about this.* (C) really is true, after all.

The paradox comes not from the truth of the premises and conclusion, but rather from the way that we can know the truth of these claims. The problem is that externalism was established on the basis of a priori philosophical argument; so, if (1) is true, then it seems that we can know it to be true a priori. And if the thesis of privileged access is true, then we can know (2) independently of empirical experience — and hence, you might think, a priori — as well. Note again: *we still do not have a paradox; it is not terribly surprising that we can know (1) and (2) a priori.*

The paradox comes in with the claim that if (1) and (2) are knowable a priori, (C) must be as well. We get this from the claim that instances of the following schema are true:

Transmission of a prioricity across logical consequence

> If A knows a priori that p, and A knows a priori that if p, then q, A is in a position to know a priori that q.

This is a plausible principle about a priori knowledge; if one knows that today is Wednesday, and knows further that if today is Wednesday then tomorrow is Thursday, it does not seem that one needs to do further empirical research to determine that tomorrow is Thursday.

Given all this, we can derive the paradoxical conclusion with the following filled-out version of Boghossian's argument:

1. If I have the concept *water*, then water exists. (Externalism)
2. I have the concept *water*.
3. (1) is knowable a priori. (A priori status of externalism)
4. (2) is knowable a priori. (Privileged access thesis)
5. It is a logical consequence of (1) and (2) that water exists.
6. A prioricity transmits across logical consequence.

C. It is knowable a priori that water exists.

Notes on the argument. We should distinguish between a priori knowledge and introspective knowledge; but this by itself does not help to resolve the paradox, since it is just as troubling that we should be able to know that water exists on the basis of introspection.

As Boghossian puts it, putting together externalism and self-knowledge, the externalist "is in a position to conclude, a priori, that water must have existed at some time. And that, we are all agreed, is not something he ought to be able to do."

The puzzle, then, is this: given that the above argument is valid, and given that it's conclusion is false, one of the premises in the argument must be rejected. Rejecting (2) is implausible, as is rejecting (5). So that gives us four premises that we can reject, and hence four ways to respond to the paradox:

1. Externalism (as stated in premise 1) is false.

2. A prioricity does not transmit across logical consequence; one can know that p and know that if p, then q without being in a position to deduce without further empirical evidence that q.

3. Externalism is true, but is only knowable a posteriori; hence premise 3 is false.

4. The privileged access thesis is false; we can only know the contents of our own thoughts a posteriori, on the basis of empirical observation.

We will consider these possibilities in turn.

2.3 Rejecting externalism

We should consider the possibility that externalism is true in some form — i.e., that it is true that intrinsic duplicates can vary with respect to certain mental properties, such as their beliefs — but that externalism does not support premise 1.

Consider, in this vein, Boghossian's (and Burge's) *dry earth scenario*, in which it appears to someone that there is a clear drinkable liquid running out of faucets etc., but in fact there is no such liquid: the appearance of such a liquid is an illusion. Would agents in such a scenario be able to have thoughts involving the concept *water*? If so, then (1) should be rejected. But it seems as though the intuitions which counted in favor of externalism count in favor of the claim that dry-earthers lack the concept *water*. To see this, just ask why we should ascribe the concept *water* rather than the concept *twater* , or XYZ, to them.

Exceptions: (i) Worlds where water once existed, but no longer does. This shows that we need to slightly complicate the conclusion of the paradox; but this doesn't matter much for the force of the paradox. (ii) Worlds where hydrogen and oxygen exists, but not water, and the agent in question has sophisticated chemical knowledge; stipulate this possibility away. We can restrict the paradox to chemically ignorant agents.

2.4 Rejecting the transmission of a prioricity across logical consequence

There are apparent counterexamples to the thesis that *whenever* one knows that p and knows that if p, then q, one also knows that q. The example of Paderewski. But these do not seem relevant to the case at hand.

2.5 Rejecting the a priori status of externalism

Maybe premise 1 is shown to be true by externalism arguments; but perhaps we should then reject the claim that it is knowable a priori, without which there is no paradox.

This may seem extremely implausible, since we did not appeal to any a posteriori principles in arguing for externalism.

But now consider the dry earth scenario again. We may know a priori that people on wet earth mean something different than people on dry earth by 'water'; but can we know a priori whether we are on wet earth or dry earth? Plausibly, the answer is, 'No.' But if this is right, then we cannot know a priori whether we mean by 'water' what those on dry earth mean by the word, or whether we mean what those on wet earth mean by it; for this reason it seems that we cannot know a priori whether we express the beliefs using this word that those on dry earth express, or whether we express the beliefs that those on wet earth express.

How does this relate to the question of whether one can know a priori that if I have

the concept *water*, then water exists?

2.6 Rejecting privileged access

Rejecting privileged access does not seem plausible; it is just obvious that we can know about what we are thinking in a way that we cannot know what other people are thinking.

It is plausible that we can know what we are thinking in some sense; but it is not plausible that we can know everything about the contents of our thoughts only by introspection. We cannot always know, for example, whether the those contents are true, or whether certain concepts have a reference.

Application of this to the wet earth/dry earth distinction. Two different views about what the meaning of 'water' would be on dry earth, and two corresponding different resolutions to the paradox.

3 The argument from slow switching

The argument from illicit a priori knowledge has more to do with Putnam's arguments for externalism, and the claim that the contents of our beliefs are in part determined by facts about our environment, than with Burge's 'social' externalism. But, as it turns out, we can also generate an argument for the incompatibilism of self-knowledge and externalism using Burgean externalism.

A brief statement of the argument may be found on pp. 124-5 of the coursepack, near the end of Boghossian's 'Content and self-knowledge.' First imagine an agent A who moves from a linguistic community x where 'arthritis' means arthritis to a distinct linguistic community y where the same word means a disease of the joints or thigh. Then we can argue as follows, using 'arthritis' to mean what it normally means in English and linguistic community x, and 'tharthritis' to express what it means in linguistic community y.

1. When A is in x, he has thoughts involving the concept *arthritis*, and that he knows that he has thoughts of this kind.
2. A could be switched into linguistic community y without knowing it. Eventually, as he comes to be a member of y, he will come to mean tharthritis, rather than arthritis, by the word 'arthritis.' This could happen without his noticing.
3. Suppose that A is switched in 2003. Then we can imagine that someone (in y) asks A in 2005 to list the concepts he had thoughts about in 2003. He responds (now using the language of y) with a long list of words, one of which is 'arthritis.' But now of course 'arthritis' means tharthritis; which means that A left at least one concept he had thoughts about off the list. So A no longer knows that he had thoughts involving the concept *arthritis*before 2003; he now thinks (falsely) that he had thoughts involving the concept *tharthritis*.
4. We can stipulate that A has a perfect memory — i.e., that he never forgets anything.
5. If someone at some time t1 knows a proposition p, and does not forget anything between t1 and t2, then he still knows p and t2.
6. So A did not know in 2003 that he had thoughts involving the concept *arthritis*.

By now the paradox should be apparent: (1) and (6) contradict each other. Since there are no true contradictions, we must again reject one of the premises. (6) follows from (3)-(5), so we cannot reject it without rejecting one of those three. There is therefore pressure to reject either externalism — the view that A's beliefs change — or the view that he knew anything about his own beliefs before the switch. The latter seems not to be plausible.

An alternative: reject the legitimacy of the 'stipulation' that A did not forget anything between the two times.