



WCP 2008 Proceedings
Vol.42 Philosophy of Mind

Agent's Knowledge and First-person Authority

Stephen Voss

Bogazici University, Turkey
voss@boun.edu.tr

ABSTRACT I propose the hypothesis that our knowledge of our own mental states derives from our knowledge of our intentions, and that our knowledge of our intentions is part of having those intentions. I enumerate various aspects of the question to be answered and various aspects of my answer. The hypothesis begins to explain various aspects of self-knowledge, such as its fallibility and its variability from one kind of mental state to another. Self-knowledge is also grounded in our common antecedent knowledge of the functionalist nature of mental states and the integrity of our mental life and above all mind's link with action in the world. The hypothesis helps pluck philosophy from the fairy-land of Lockean and Kantian inner sense.

I'd like to propose a hypothesis to explain our knowledge of our own mental states: that it derives from our knowledge of our intentions to do things. I'll note significant aspects of the question and significant elements of my answer. I'll concentrate on one type of mental state—belief—but I'll indicate how to think about other types.

Here are three questions that arise immediately:

Question 1. Most importantly, what explains the fact that we are often in a better position to know our own mental states than other people are?

Question 2. What explains the obvious limits to our authority? Why aren't we infallible?

Question 3. Why do we know more about some kinds of states than others? People are not normally in doubt about what they *believe*. When we're in doubt, it's not about whether we believe that p, it's about p. By contrast, people are often in doubt about whether they regret something or are proud of something, and very much in doubt about whether they love someone or whether they hate someone.

*

I know that I believe that there is global warming. It would be wise to ask at the outset what sort of thing this belief is. I think that the right answer is functionalist, and I think it's the first step in answering our questions.

Answer 1. Functionalism is true.

A belief, then, is a cluster of dispositions that certain kinds of stimuli and mental states tend to cause: dispositions to act in certain ways and to have certain other mental states.

Functionalism sharpens our questions before it begins to answer them. For a belief is an indefinitely large and complex cluster of dispositions, impossible in principle to specify fully. Functionalism makes 'S believes that p' literally undefinable. Given that no one will ever be able to specify the dispositions that constitute belief that p, and given that all but a handful of them will remain forever unrealized, how can a person know that a state he is in satisfies the conditions for being a belief that p, and is not instead a belief in something very similar, or an attitude very similar to belief? How could anyone ever possess sufficient insight into an innumerable host of *unrealized* dispositions to learn that it is precisely a belief that the earth is warming? Yet no one ever says 'As

far as I can tell, I believe that the earth is warming ... but let me get back to you on that'. We must ask a new question.

Question 4. What explains our knowledge about states that we ourselves understand only as innumerable hosts of largely unrealized dispositions?

*

One thing that belief that *p* disposes me to do is to affirm that *p* when I want to tell the truth about whether *p*. Let us begin here.

Believing that *p* is much more than simply tending to affirm that *p* when one wants to tell the truth. So it is possible to affirm sincerely that *p* even if one does not believe that *p*. Thus I might affirm that airplanes are safe although you can't get me into one. Here my assent comes apart from my nonverbal behavior. In such a case I might know that I am sincerely affirming that *p* and *not* know whether I believe that *p*.

But this is unusual. Such a lack of integrity among the items in a functionalist bundle constituting a mental state is the exception to the rule. This is something we can each learn by observing the degree of coherence in our own behavior.

Answer 2. A subject's mental life frequently contains coherent bundles of dispositions that qualify as mental states—coherent in the sense that, for example, a subject's verbal behavior and nonverbal behavior suggest the same answer to questions about the subject's mental state.

And this congruence is part of the explanation of our authority. It will help answer Question 4.

This fact suggests a deeper question.

Question 5. Why does a subject's mental life so often contain coherent bundles of dispositions?

Still, we must not exaggerate the usual level of coherence. As James Joyce's novels remind us, our mental life contains many fragments as well, which do not cohere much with our other dispositions. Another

fact will help explain the limits of our knowledge of our mental states. That fact in turn raises a new question.

Answer 3. A subject's mental life often contains fragments not coherent with his other dispositions.

Question 6. Why is that?

If Answers 2 and 3 help answer our earlier questions, answers to Questions 5 and 6 would help in deeper ways. I don't know the answers to them.

*

The tradition offers two paradigms of epistemic authority. One is that of the perceiver. The other is that of the agent. Given the failures of the perceptual model for first-person authority, it's time to investigate the possibilities of the opposing model.

*

I believe that in formulating his intentions, an agent necessarily knows what they are. First, he knows what he intends to do: he knows his *actual* intentions. Secondly, he knows his *conditional* intentions: what he intends to do in case certain conditions are realized. I shall call such knowledge of one's intentions *agent's knowledge*.

Stuart Hampshire asserts in *Thought and Action* that if someone acts intentionally or intends—actually or conditionally—to act, it *follows* that he knows what he intends to do. His reason is that knowing what you intend to do is part of intending to do it. Only that knowledge, he says, gives you opportunity for choice and with it full responsibility for what you do.

This seems right. It then answers Question 1 for one specific case: our knowledge of our intentions.

Answer 4. A subject knows his actual and conditional intentions, because such knowledge is part of intending to do something.

Below I will suggest how Answer 4 not only answers Question 1 in the case of intentions, but is an essential aspect of the answer to Question 1 for *every* type of mental state.

Answer 4 leaves a deeper problem: how it is that a subject can act voluntarily at all, given that this involves knowing what one intends to do? How is it that a being can be an agent?

Question 7. What explains a subject's capacity for voluntary action, given that this requires the capacity to know what one intends to do?

*

If S believes that p then S tends to affirm that p when S wants to tell the truth about whether p. A subject's sincere affirmations provide a good test of that subject's beliefs.

Now suppose that I have the following conditional intention: I intend to affirm that airplanes are safe when I want to tell the truth about whether they are. I will then know that I have that intention. True, given the possibility of fragmentation in my mental life, I may not satisfy enough of the other conditions for belief, and therefore it may be that even if I *believe* that I believe that airplanes are safe, I don't *know* I believe it. Still, my knowledge that my mental life has a good deal of integrity gives me *some* reason to believe that I do believe that airplanes are safe. Is there anything I can do to strengthen this reason and render my belief authoritative?

*

Yes: I can review my non-verbal intentions to act. Would I *act* as though I believed that airplanes are safe? I've just sincerely affirmed it. Now, wondering whether I really believe it, I ask myself whether I would actually get on that plane. Picturing myself at the door of the plane, what do I find myself intending to do?

If my review of my other intentions shows that they cohere with my affirmation, I can claim to know that I believe that p. If it shows a

divergence, the claim is called into question. I may have to become more clear about the correct account of belief.

*

Belief is more even than a set of tendencies toward voluntary action. A subject might then conceivably pass all of the tests I've just spoken of and still not believe. This possibility sets a limit to the authority a subject can possess concerning his beliefs.

Answer 5. A subject is capable of authoritative belief about whether he believes that *p* insofar as he is capable of extrapolating rationally from the inclinations to act that are part of the bundle constituting belief that *p* to the bundle itself.

The risk in this extrapolation helps explain why authority is always a matter of degree, falling short of infallibility.

*

We can easily extend these considerations to mental states other than belief. In the case of belief, we found one central but not decisive indicator: a tendency to assert what one believes when one is sincere. Desire has a similarly central indicator: a tendency to do actions when one believes that they will yield what one desires. The remaining basis for clarifying what one desires is the remaining tendencies to voluntary action in the bundle constituting desire. Epistemic access to all these tendencies is provided by agent's knowledge: knowledge of one's own intentions. The limits to our knowledge of our desires are grounded just as the limits of our beliefs are.

Other kinds of mental states typically lack such a central indicator. That lack will help explain why we are less authoritative about those cases than we are about our beliefs and our desires.

A new difficulty may arise for conceptually complex kinds of mental states. A subject may be unsure of the correct functionalist account of some kind of mental state. I may know much about how I intend to act

around the lady and yet not be sure whether I love her. Agent's knowledge makes self-knowledge possible, but it doesn't always make it easy.