Smithies on Self-Knowledge of Beliefs

Brie Gertler, University of Virginia

*This is a penultimate draft. Please refer to the published version.*

In *The Epistemic Role of Consciousness*, Declan Smithies develops a systematic, cohesive account of beliefs, phenomenal consciousness, epistemic justification, and rationality. The arguments for the various elements of the account are extremely rigorous, and the account is impressively broad in scope. This commentary will focus on Smithies’ views about self-knowledge. Specifically, I will examine his case for the striking thesis that rational thinkers will know all their beliefs. I call this *the ubiquity of self-knowledge* thesis. Smithies’ case for this thesis is an important pillar of his larger project, as it bears on the nature of justification and our ability to fulfill the requirements of rationality.

Section 1 outlines Smithies’ argument for the ubiquity of self-knowledge. Section 2 sets the stage for a detailed explication of his view by sketching an initial objection to that argument. Sections 3 and 4 respond to that objection on behalf of Smithies’ account, elucidating the distinctive positions on belief and justification that enable this response. Sections 5-7 raise some questions, namely: whether Smithies’ views about self-knowledge of beliefs provide a genuine competitor to other philosophers’ views on that issue; whether on his account justification is truly accessible to the thinker; how his account explains self-knowledge; and what his account implies about the situation of actual persons, who are not ideally rational.

1. Smithies’ argument for the ubiquity of self-knowledge

Moore-paradoxical beliefs (in the omissive form) are beliefs with the content *p but I don’t believe that p*. Such beliefs are always false, since belief in the first conjunct entails the falsity of the second conjunct. This means that Moore-paradoxical beliefs are self-defeating in virtue of their structure: the content of a Moore-paradoxical belief is falsified by being believed.

Smithies takes it as a datum that Moore-paradoxical beliefs are irrational and argues that the best explanation of their irrationality is the thesis I call *the ubiquity of self-knowledge*. Here is that argument:

1. Moore-paradoxical beliefs are always irrational.

“it can never be rational to form false beliefs about what you believe” (172)[[1]](#footnote-2).

1. The best explanation of (i) is that a rational agent will know all of her beliefs.

“Rationality requires knowing what you believe, since otherwise you’re liable to fall into an irrational Moorean predicament” (174).

Therefore,

*The ubiquity of self-knowledge*: A rational thinker will know all of her beliefs.

2. An objection to premise (i)

It may appear that Moore-paradoxical beliefs are always irrational. After all, a rational thinker will be able to recognize that beliefs of the form *p but I don’t believe that p* are self-defeating. So rational thinkers—who, let’s assume, will try to avoid false beliefs—will try to avoid Moore-paradoxical beliefs. However, there is no guarantee that they will succeed in this. For there is no guarantee that a rational thinker will be able to recognize that a particular belief is a Moore-paradoxical belief, or so I will argue.

To make this case, I’ll add a twist to the following well-known example from Christopher Peacocke:

Someone may judge that undergraduate degrees from countries other than her own are of an equal standard to her own, and excellent reasons may be operative in her assertions to that effect. All the same, it may be quite clear, in decisions she makes on hiring, or in making recommendations, that she does not really have this belief at all. (Peacocke 1999, 90)

Let’s call the subject in this example *Molly*. It is clear from this passage that, on Peacocke’s view, belief is at least partly constituted by the dispositions to reason and act that are manifested in Molly’s evaluations of applications and related decisions (compare Schwitzgebel 2002).

Here is my twist on Peacocke’s example. Suppose that Molly understands that when it comes to issues that tend to trigger biases, such as the value of foreign degrees, our conscious judgments often diverge from our dispositional beliefs. Reflecting on that fact prompts her to return to a set of applications that she’s already ranked, with an eye towards discovering whether her dispositions match her judgments—that is, whether she is disposed to act and reason as if foreign degrees are as valuable as domestic degrees. Taking the relevant factors into account, she finds that her rankings do favor those with domestic degrees. On the basis of that evidence, together with her background recognition that biases often affect beliefs like these, she comes to believe *I don’t believe that foreign degrees are as valuable as domestic degrees*. Given her evidence, that belief is rational.

We may further suppose that, although this meta-belief is rationalized by evidence, it is false. A false yet justified belief of this kind could arise in a number of ways. Here is one. Unbeknownst to Molly, a nefarious agent interfered with the applications after her initial assessment so that even the most careful and comprehensive review would lead a rational thinker to conclude *I don’t believe that foreign degrees are as valuable as domestic degrees*.

In this case, it’s hard to deny that Molly believes, on rational grounds, both that (a) foreign degrees are as valuable as domestic degrees, and that (b) she does not believe that these degrees are equally valuable. Regarding (a): she is disposed to act and reason as if foreign and domestic degrees are equally valuable; those dispositions derive from her earlier deliberations about their relative value; and she is disposed, when she considers the question, to judge that these degrees are equally valuable. Regarding (b): given the evidence manufactured by the nefarious agent, she is entirely rational in believing that she is not disposed to act and reason as if these degrees are equally valuable. And she has no reason to think that the dispositions that shaped her rankings—which (she reasonably believes) survived her previous explicit commitments to equity—would change simply upon her recognizing them.

In sum, it appears that Molly rationally believes that *p*, and rationally believes that she does not believe that *p*. She is disposed, on rational grounds, to act and reason as if *p*; and she is also disposed, on rational grounds, to act and reason as if she does not believe that *p*.

Moreover, Molly may rationally believe the conjunction *foreign degrees are as valuable as domestic degrees, but I don’t believe that they are*. This conjunctive belief is manifested in her reasoning and actions as she commits herself to eradicating what she takes to be her tendency to favor applicants with domestic degrees, deliberates about how to accomplish this, and enacts the plan she has devised.

In this example, the subject makes a false self-ascription: she mistakenly thinks that she believes that foreign degrees are inferior to domestic degrees. I have argued that because that error is based on good (albeit misleading) evidence, it does not represent a failure of rationality. So contrary to premise (i) of Smithies’ argument for the ubiquity of self-knowledge, Moore-paradoxical beliefs can be rational.

3. Smithies’ conception of belief

In this section and the next, I elucidate key features of Smithies’ account by explaining the account’s resources for handling the objection just given.

Smithies allows that situations like Molly’s, in which one has evidence for *p* while also having evidence for *I don’t believe that p*, appear to be possible. And he recognizes that thismakes it seem as if a Moore-paradoxical belief—and the associated lack of self-knowledge—could be rational. To preserve his conviction that Moore-paradoxical beliefs are always irrational, he needs to block such examples. He does this by combining a somewhat idiosyncratic conception of belief with an expansive notion of epistemic justification.

On Smithies’ view, beliefs “are individuated by their dispositions to cause phenomenally conscious judgments: to believe that *p* is to be disposed to judge that *p* when you consciously entertain whether *p*” (175-6). A belief that *p* is not even partly constituted by dispositions to act and reason (though dispositions to judge will sometimes generate such dispositions). So regardless of whether Molly’s ranking of applications favors domestic degrees, on Smithies’ view the fact that she’s disposed to judge that *foreign and domestic degrees are equally valuable* means that she believes this. Similarly, the fact that she’s not disposed to judge that *foreign degrees are inferior to domestic degrees* means that she does not believe that.

Smithies’ view does not collapse the distinction between judgment and belief. He thinks that it’s possible to judge that *p* without believing that *p*, and vice versa, since judgments do not always manifest dispositions. He gives an example of someone who is primed with an image of the Sydney Opera House just before being asked to name the capital of Australia, and responds to this request by judging *Sydney is the capital of Australia*. That judgment may be a momentary lapse, in that the person may well know—and be generally disposed to judge—that Canberra is the capital. In such cases, “your judgment fails to express what you believe” (184).

Smithies’ conception of belief implies that Molly’s review of her application rankings does not justify her in denying that she believes *foreign and domestic degrees are equally valuable*. On his conception, how one is disposed to act and reason is not directly relevant to what one believes. So Molly’s review of her rankings sheds no light on what she believes. After all, her purpose in undertaking that review is precisely to determine whether her dispositions to act and reason *diverge from* her judgments (or dispositions thereto).[[2]](#footnote-3) If beliefs are dispositions to judge, then Molly is not justified in her Moore-paradoxical belief. So her situation is not a counter-example to premise (i) of the argument I sketched earlier, which says “it can never be rational to form false beliefs about what you believe” (172).

Of course, this does not imply that premise (i) is true. But let’s leave that question and turn to examining Smithies’ account of self-knowledge. If rational thinkers know all of their beliefs, as the ubiquity of self-knowledge thesis says, how is this knowledge achieved?

4. The Simple Theory of Introspection

As we just saw, judgments can diverge from beliefs even on Smithies’ conception of beliefs: judging that *p* does not guarantee that one is disposed to judge that *p*. Still, judging that *p* may be *defeasible* evidence that one believes that *p*. But the availability of defeasible evidence would not explain the ubiquity of self-knowledge. Smithies thinks that, since “[r]ationality requires that you know your beliefs”, a rational thinker who believes that *p* will have *infallible* evidence that they believe that *p* (186). Since the inference from a judgment to the corresponding belief is not always truth-preserving, the judgment that *p* is not infallible evidence that one believes that *p*. So we must look elsewhere for an account of how rational thinkers know their beliefs.

Smithies’ account of how rational thinkers know their beliefs partly consists in his “simple theory” of introspection.

The simple theory of introspection … says that when you have an introspective reason to believe that you’re in a certain kind of mental state, you have that reason just by virtue of being in that mental state. There is no further requirement that you have some inner perception that you’re in that mental state, or that you have reason to believe the premises of a noncircular argument for the conclusion that you’re in that mental state. The mere fact that you’re in that mental state is sufficient to give you an introspective reason to believe that you’re in that mental state. (154-5)

According to this theory, being disposed to judge that *p* suffices for having a reason to believe that you’re disposed to judge that *p*. Importantly, for Smithies “introspection” does not refer to a type of observation. A disposition to judge need not be—arguably could not be—observed directly. And it need not be inferred from what is directly observed, such as a phenomenally conscious judgment that manifests it. Instead, simply having the disposition gives you reason to believe that you have it.

This idea, that having a particular disposition can provide non-inferential reason to believe that one does, is somewhat curious. As a general matter, we seem to know dispositions by inference from evidence about their manifestations or bases. We know that the vase is fragile either by inference from observing similar objects break (in circumstances in which a non-fragile object would remain intact) or by inference from its constitution (e.g. that it is fine porcelain). When it comes to psychological dispositions, self-knowledge is also inferential, and usually consists in inference from real or simulated manifestations of the disposition. I know that I’m disposed to feel joy when I see a puppy by inference from past manifestations of this disposition, or by inference from the phenomenology I experience when imagining puppies.[[3]](#footnote-4)

Smithies embraces the simple theory because he sees it as needed to explain how rational thinkers are able to meet the requirements of rationality, such as the requirement that one avoid Moore-paradoxical beliefs. (Rationality also requires that we avoid epistemic akrasia, e.g., believing that *p* while simultaneously believing that one lacks justification for believing that *p*.) Rational thinkers can avoid Moore-paradoxical beliefs only if beliefs are “luminous”, where a class of mental states is luminous *iff* “you’re always in a position to know by introspection whether or not you’re in those mental states” (26). The claim that beliefs are luminous is, in effect, the simple theory of introspection applied to beliefs. To say that beliefs are luminous is to say that believing that *p* suffices for reason to believe that you believe that *p* and—of direct relevance for the rationality of Moore-paradoxical beliefs—reason not to deny that you believe that *p*. The simple theory of introspection secures luminosity and thereby ensures that rational thinkers can avoid Moore-paradoxical beliefs.

The next three sections raise a worry and two questions about Smithies’ picture of self-knowledge of beliefs. The worry is that this picture doesn’t seem to apply to “beliefs” in the standard sense of that term. The first question concerns how we actually achieve what Smithies takes rationality to require, that is, how it is that we are in a position to know our beliefs and to know whether they are epistemically justified. The second question concerns whether Smithies’ ambitious vision of what rationality requires sheds light on the epistemic situation of actual human beings.

5. Self-knowledge of “belief”

The ubiquity of self-knowledge thesis, adapted to Smithies’ conception of beliefs, amounts to the thesis that we always know our dispositions to judge. That is: if a rational agent is disposed to judge that *p*, then she will know that she is; if she is not disposed to judge that *p*, then she will know that she is not.

This thesis may well be of some interest. But I don’t think it addresses what philosophers have meant by “self-knowledge of beliefs”. Philosophers who consider how one knows one’s own beliefs are generally interested in one of two kinds of attitude.

1. The kind of attitudes about *p* described in Section 2 above, which are partly constituted by dispositions to act and reason as if *p.*
2. Intrinsically normative commitments as to *p*,which consist in an exercise of cognitive agency and for which the thinker bears responsibility.

As we saw earlier, Smithies’ conception of belief diverges from (1). And as far as I can tell, Smithies’ conception also diverges from (2), since he does not envision belief as freighted with agency and normativity.

This last point deserves elaboration. Smithies may well see *rationality* as intrinsically normative—that is, of intrinsic and not simply instrumental value. (See his discussion of the epistemic “ought”, pp. 271-5.) But although we may display epistemic excellence when our beliefs are properly responsive to reasons, Smithies does not define beliefs—of the sort he has in mind—in terms of that responsiveness. By contrast, philosophers who argue that what is truly distinctive about self-knowledge derives from our capacity to have type (2) attitudes sharply distinguish beliefs from what they regard as mere (“brute”) dispositions.

One may be disposed to judge that *p* without being (1) disposed to act and reason as if *p*, or (2) willing to regard *p* as a commitment (or to avow it). So it’s not clear that Smithies’ view about self-knowledge of beliefs addresses the questions of concern to most philosophers working on that issue. Of course, Smithies’ view may nevertheless have important implications for philosophy of mind and epistemology. And it is a direct competitor to various views on other issues, such as the nature of introspective justification.

6. Accessibility and the phenomenal

Recall that for Smithies, beliefs “are individuated by their dispositions to cause phenomenally conscious judgments: to believe that *p* is to be disposed to judge that *p* when you consciously entertain whether *p*” (175-6). He takes this to imply that beliefs are “phenomenally individuated”. This is a very liberal standard for phenomenal individuation. It means that purely dispositional differences can suffice for “phenomenal” differences. Since the relevant dispositions may never be manifested, the result is that a purely counterfactual difference between two thinkers, as to what phenomenal state they would experience if they were to consider whether *p*, constitutes a phenomenal difference. (I return to this point below.)

The fact that beliefs are phenomenally individuated means that they can contribute to the thinker’s epistemic justification, compatibly with Smithies’ *Phenomenal Mentalism*.

*Phenomenal Mentalism*: Necessarily, which propositions you have epistemic justification to believe at any given time is determined solely by your phenomenally individuated mental states at that time. (25)

At first glance, Phenomenal Mentalism may appear to express what Williamson has called the “phenomenal conception of evidence”: the idea that states with phenomenal character exhaust the evidential sources of justification (Williamson 2000: 173). However, Smithies’ standard for phenomenal individuation yields a much broader category of evidential states. Suppose that there is someone phenomenally indiscernible from you in the following sense: at every moment throughout both of your lifetimes, you and that person have precisely the same phenomenal states (states with phenomenal character). Phenomenal Mentalism does not imply that you and that person would be indiscernible with respect to epistemic justification. For there may be counterfactual situations in which you and that person would have different phenomenal states: e.g., if you both had considered whether *p*, you would have consciously judged that *p* while this other person would not. Smithies’ construal of “phenomenal individuation”, when combined with his Phenomenal Mentalism, yields the result that such purely counterfactual differences can constitute differences in epistemic justification.

It is therefore striking that part of the support for Phenomenal Mentalism is that it “explains why you and your phenomenal duplicates in skeptical scenarios have epistemic justification to believe the same propositions to the same degree” (25). This claim employs a conception of phenomenal duplicates that is unusually restrictive, corresponding to the unusually liberal construal of phenomenal individuation. On this restrictive conception, phenomenal duplicates must not only share all phenomenal states, but must also be phenomenally identical across a range of counterfactual situations. A person whose phenomenal life precisely matched yours throughout both of your lifetimes would not be your phenomenal duplicate so long as there was some counterfactual situation, never encountered, in which you and that person would have had different experiences. On the more standard conception, all that’s required for being phenomenal duplicates is indiscernibility with respect to actual phenomenal states.

The issue is not merely verbal. Allowing that purely counterfactual differences in phenomenal states can make a difference to epistemic justification is crucial to Smithies’ claim that believing that *p* provides a reason to believe that you believe that *p*. For a belief, as a *disposition to* phenomenally conscious judgments, may have no effect on the believer’s actual phenomenology. On a phenomenal conception of evidence (in Williamson’s sense), indiscernibility with respect to *actual* phenomenal states entails indiscernibility with respect to evidence. That view is inspired by the idea that sources of justification must be internal in both of the familiar senses: mental and accessible. Phenomenal states are clearly mental (they may be the paradigm of mentality). And they are taken to be particularly accessible, either because they are introspectible or simply because they constitute the qualitative character of experience.

In effect, Smithies is proposing that justification need not supervene on actual phenomenal states in order to satisfy the mentality and accessibility requirements on justification. That he takes his proposal to satisfy these requirements is clear from his choice of the labels “Mentalism” and “Accessibilism” to refer to core elements of his view.

But does Smithies’ proposal satisfy the spirit of these requirements? I have some doubts, especially as regards accessibility. Here is how Smithies understands accessibility.

*Accessibilism*: Epistemic justification is *luminously accessible* in the sense that, necessarily, you’re always in a position to know which propositions you have epistemic justification to believe at any given time. (28)

Accessibilism follows from two claims already mentioned. First, the Phenomenal Mentalism thesis, which says that epistemic justification is determined solely by phenomenally individuated states (which includes beliefs); and second, the claim that phenomenally individuated states “are ‘luminous’ in the sense that you’re always in a position to know by introspection whether or not you’re in those mental states” (26).

Although these claims secure Accessibilism, I have trouble seeing exactly *how* it is that we are able to know our phenomenally individuated states and, hence, our epistemic justification. Smithies says:

The best explanation of accessibilism is that your evidence is exhausted by introspectively luminous facts about your mental states. But only phenomenally individuated facts about your mental states are introspectively luminous in the requisite way. (28)

It remains unclear to me how it is that being phenomenally individuated makes a mental state (or facts about it) luminous. Suppose that I believe that *p*—that is, I am disposed to consciously judge that *p*—while a person who has all and only the same phenomenal states as I do, throughout both our lifetimes, does not believe that *p*. (In this scenario, neither of us ever entertain the question whether *p*, so our respective dispositions to judgments as to whether *p* are never manifested.)On Smithies’ view, I am epistemically justified in believing that I believe that *p* (purely in virtue of having the first-order belief), whereas the person whose phenomenology perfectly matches my own is not justified in believing that she believes that *p*. But without a difference in actual phenomenology, it’s not clear how we can respond to our epistemic situations in ways that reflect the difference between them: how I can know that I’m epistemically justified in believing that I believe that *p*, and how she can know that she is not. So it’s not clear how Smithies’ picture secures the truth of Accessibilism.

This is not to say that phenomenology is the only possible source of knowledge about justification. But I would appreciate a fuller picture of how we manage to be in a position to know our epistemic justification if that can vary independently of actual phenomenology. In particular, I would like to know how the appeal to phenomenal individuation is intended to help explain the accessibility of justification.

7. The non-ideally rational

To my mind, Smithies’ intriguing proposal is, at its most basic, a proposal about the epistemic situation of a thinker who meets certain rational requirements: who avoids Moore-paradoxical beliefs, who is never epistemically akratic, etc. For such a thinker, having a belief will equip them to know that they do; epistemic justification will always be luminously accessible; etc. The pieces of this picture fit together very neatly.

But it is not altogether clear how this picture relates to the situation of actual human beings. For Smithies, “ideal rationality is a matter of proportioning your beliefs to the evidence” (28). Of course, we are somewhat rational, in that we sometimes proportion our beliefs (roughly) to our evidence. But our beliefs often diverge from what our evidence supports. And we are probably incapable of consistently and precisely proportioning our beliefs to our evidence. Even if we decided to put aside all other goals and strive single-mindedly for ideal rationality, we would not eliminate all divergences between our beliefs and our evidence. If that is correct, Smithies’ proposal does not describe an epistemic situation in which we could find ourselves.

So how does that proposal bear on our cognitive and epistemic situation? One possibility is that Smithies’ intention is to flesh out an ideal of rationality for which we should strive. That interpretation does not seem promising to me, however. Smithies’ claims about what rationality requires, which anchor his account, are supported by their intuitive force rather than by arguments. The claims that he supports by argument, such as Phenomenal Mentalism, are high-level descriptions that do not provide guidance for navigating our cognitive lives.

Here is another way to think about this issue. Suppose that Smithies is right that anyone who believes that *p* has reason to believe *I believe that* *p*. This means that, if I were ideally rational and believed that *p*, then—at least if I considered whether I believed that *p*—I would believe *I believe that p*. However, none of us is ideally rational: we do not always proportion our beliefs to our reasons. So how does Smithies’ account bear on non-ideally rational thinkers like us? How do we *use* the reasons that Smithies says we possess? Is there any guarantee that we can form the appropriate beliefs in response to those reasons? If there is such a guarantee, what grounds it? And if there is no such guarantee, then in what sense can we be said to possess those reasons?

**Conclusion**

Smithies has offered a remarkably systematic picture of the nature of and relations among belief, epistemic justification, phenomenal consciousness, and self-awareness. The bold originality of his approach to basic issues in epistemology has the salutary effect of laying bare and challenging some existing orthodoxies. I would like to know more about the cognitive processes by which we come to know our beliefs and our justification. More generally, I would welcome an explanation of how this picture relates to the epistemic situation of flawed, limited creatures like us.[[4]](#footnote-5)

Works cited

Gilbert, D. et. al. (2009) “The Surprising Power of Neighborly Advice”. *Science* 323: 1617-19.

Peacocke, C. (1999) *Being Known*. Oxford: Oxford University Press.

Schwitzgebel, E. (2002) “A Phenomenal, Dispositional Account of Belief”. *Noûs* 36: 249-75.

Smithies, D. (2019) *The Epistemic Significance of Consciousness*. Oxford: Oxford University Press.

Williamson, T. (2000) *Knowledge and its Limits*. Oxford: Oxford University Press.

1. Page numbers without an author name refer to Smithies 2019. [↑](#footnote-ref-2)
2. We can stipulate that in this case Molly’s judgments manifest her disposition to judge. This may simply make explicit what was implicit in Peacocke’s original description of the case; more to the point, this stipulation does not beg any questions or threaten the intuitive force of the example. [↑](#footnote-ref-3)
3. Self-knowledge of one’s own dispositions may be achieved through inference from other kinds of evidence as well. Some social psychologists argue that, when it comes to some dispositions—e.g., how we will respond or feel in certain circumstances—our friends’ responses in similar circumstances provide powerful evidence (Gilbert et. al. 2009). [↑](#footnote-ref-4)
4. I am indebted to Declan Smithies for valuable discussion of many of the points in this commentary. [↑](#footnote-ref-5)