Philosophers' Imprint

VOLUME 17, NO. 16
AUGUST 2017

Self-Blindness and Self-Knowledge

Matthew Parrott

King's College London

Abstract: Many philosophers hold constitutive theories of self-knowledge in the sense that they think either that a person's psychological states depend upon her having true beliefs about them, or that a person's believing that she is in a particular psychological state depends upon her actually being in that state. One way to support this type of view can be found in Shoemaker's well-known argument that an absurd condition, which he calls "self-blindness", would be possible if a subject's psychological states and her higher-order beliefs about them were wholly distinct existences. A second reason to endorse a constitutive theory is the widespread conviction that first-person access is epistemically special. In this essay, I shall argue that even if self-blindness is impossible, the best explanation for this does not deny that a person's psychological states are wholly distinct from her beliefs about them. I shall then attempt to account for the epistemic distinctiveness of first-person access on the basis of fundamental features of rational cognition. One advantage of this account over constitutive theories of self-knowledge is that it is better placed to explain our fallibility and ignorance.

In order for me to know about another person's psychological states, I must observe her in some way. I have to see what she is doing or listen to what she is saying before I can know what is on her mind. Because observation is, either directly or indirectly, necessary for my becoming aware of another person's mind, I can be, and sometimes am, mistaken about what other people are thinking. Notably, it seems possible for me to be wrong about what another person is thinking or feeling even in cases where my beliefs about the other person's mind are fully justified. The possibility of this kind of perception-based error suggests that another person's psychological state is independent from my beliefs about it. In Hume's words, the two seem to be distinct existences.

Things are different in my own case. I do not normally rely on sensory perception in order to know what I think, or want, or intend.

Instead, it seems that I have a special kind of epistemic access to my own psychological states. This first-person access is unavailable to other people. They can never know about my psychological states in the special way that I typically do. However, having this special mode of access makes it much less clear whether my psychological states are wholly distinct from my beliefs about them. Indeed, many philosophers think the two are not wholly distinct existences.[1] For example, Matthew Boyle denies that "being in a given mental state *M* and believing oneself to be in *M* are two distinct psychological conditions" (2011, pg. 235). Instead, Boyle thinks the two are simply different aspects, or "ways of conceiving", the very same psychological state. Thus, on Boyle's view, the relation between a subject's psychological state and her higher-order belief about it is identity.

In recent years, several philosophers have explicitly defended "constitutive theories" of self-knowledge. The majority of these theorists deny that a subject's psychological states are wholly distinct from her own higher-order beliefs about them.[2] Instead, they maintain either that part of what it is to be in a psychological state *M* is to believe that one is in *M*, or that part of what it is to believe that one is in *M* is to actually be in *M*.[3]

Yet there are different ways in which a constitutive theorist can think of the underlying relation that holds between a subject's psychological states and her higher-order beliefs about them. Like Boyle, one could think of the relation as identity; that is, one could think that my believing that *P just is* my believing that I believe that *P*. Alternatively, one could think of the relation as a type of essential mereological relation. For example, one could think that my higher-order belief that I believe that *P* is an *essential* part of what it is to believe that *P* and so not wholly distinct. Or, following Sydney Shoemaker, one could think that the belief that I believe that *P* "has the belief that p as an essential part", such that "its possession cannot survive the loss of the belief that p" (2009, pg. 42).[4]

Regardless, what these different constitutive theories have in common is their opposition to the natural thought that being in a psychological state is one thing and having a belief about it is another thing. I propose to capture this thought by negating two claims:

> *Distinct Existence Thesis:* For any subject *a* and psychological state *M*: (i) it is not the case that part of what it is for *a* to be in *M* is for *a* to believe (first-personally) that *a* is in

1. I intend 'wholly distinct' to mean that there are no necessary connections between the two entities. As Wilson (2010) shows, there are weaker senses of 'distinct' which allow for 'distinct existences' to stand in certain kinds of necessary relations to one another.

2. In addition to Boyle, constitutive theories are proposed by Bilgrami (2006, 2010), Coliva (2008), Heal (2001), Horgan and Kriegel (2007), Rödl (2007), Shoemaker (1994; 2012), and Zimmerman (2006). An important exception to what I say in this essay is Crispin Wright, who in various papers (1992, 1998, 2012) has defended a kind of anti-realist, "deflationary", account of self-knowledge that he sometimes refers to as a "constitutive" theory (although in recent work he concedes that this label may be misleading). Because Wright's view concerns constitutive conditions on the way we *treat* avowals of psychological states, it is not committed to denying that a subject's psychological states are wholly distinct from her higher-order beliefs about them. It therefore need not be opposed to the thesis I defend in this essay.

3. Rayo (2009) calls "part of what it is to be" claims like these "semi-identities" and argues that they entail sentences with trivial truth conditions. So, if either of these semi-identities were true for psychological states and higher-order beliefs about them, then we should expect that sentences like 'If I believe *P*, then I believe that I believe *P*' will have trivial truth conditions (cf. Rayo 2013). This suggests, as Boghossian (1989) argued, that constitutive theories make self-knowledge neither a cognitive nor an epistemic achievement (however, see Rayo [2009] for an argument that acquiring knowledge of trivial claims can be a cognitive achievement).

4. In his (1994), Shoemaker spells out the relevant relation in terms of the realizers of psychological states and of higher-order beliefs. In fact, he presents two proposals. First, he claims that *a*'s being in *M* and *a*'s believing that *a* is in *M* might have different core realizations but "their total realizations overlap in a certain way" (pg. 288). Second, he proposes that "it might be that they have the same core realization and that the total realization of the first-order state is a proper part of the total realization of the first-person belief that one has it" (pg. 288). It is worth noting that since Shoemaker clearly intends for the relation between one's psychological states and one's true higher-order beliefs about them to hold with necessity, he can't be thinking of either overlap or proper parthood as contingent relations, which is perhaps why he frames his proposals in terms of the "essence" of psychological states.

> *M*, and (ii) it is not the case that part of what it is for *a* to believe (first-personally) that *a* is in *M* is for *a* to be in *M*.[5]

Constitutive theories of self-knowledge, as I shall understand them, either explicitly deny or are committed to rejecting the *Distinct Existence Thesis*.

Rejecting this thesis amounts to endorsing the idea that part of *what it is to be* in a certain kind of psychological state is to be in another kind of psychological state. Thus, constitutive theories are committed to some type of modal truth. This is because the "part of what it is to be" operator has modal implications.[6] Specifically, the proposition negated in (i) entails the corresponding modal statement: *necessarily* (if *a* is in *M*, then *a* believes (first-personally) that *a* is in *M*). Similarly, the proposition denied in (ii) entails the corresponding modal statement: *necessarily* (if *a* believes [first-personally] that *a* is in *M*, then *a* is in *M*). It follows that someone who rejects both conjuncts of the *Distinct Existence Thesis* would commit herself to the kind of necessary bi-conditional one finds explicitly endorsed by several constitutive theorists (e.g., Bilgrami 2006; Coliva 2012). By contrast, someone who accepts the *Distinct Existence Thesis* is committed to denying both of these modal claims and maintaining instead that any relation between *a*'s belief that she is *M* and *a*'s being in *M* is contingent.

There are a couple of simple considerations that seem to speak in favor of the *Distinct Existence Thesis*. First there is Hume's doctrine that, since all psychological states are distinguishable in thought, "they may exist separately, and have no need of anything else to support their existence".[7] *Prima facie*, Hume's remark seems plausible. It seems that we can imagine a world in which a person believes that *P* without believing that she believes that *P* and also a world in which a person believes that she believes that *P* but actually does not. But if the *Distinct Existence Thesis* were false, these imagined scenarios would be impossible. Naturally, a contemporary philosopher might insist that Hume illegitimately presupposes that our imagination reliably indicates metaphysical possibility. But, even if we reject this presupposition, we might nonetheless think that being able to imagine these sorts of cases offers us at least some *prima facie* reason for thinking they are possible. After all, we might note that at least in paradigm cases where part of what it is to be F is to be G, a world in which an object is F but not G seems to be completely unimaginable. For example, it is extremely difficult to imagine a scarlet object that is not also red.

Secondly, it seems that we can be both ignorant of and mistaken about our own psychological states. We sometimes believe or desire something without believing that we do, and we are sometimes wrong about our own beliefs and desires (Schwitzgebel, 2008; Snowdon 2012). The combination of ignorance and error in this domain suggests that the *Distinct Existence Thesis* is true. More precisely, the possibility of ignorance suggests that clause (i) of the thesis is true, and the possibility of error suggests that clause (ii) is true.

Someone who is resistant to the *Distinct Existence Thesis* might think that one or both of these possibilities are exceedingly rare. That

5. Constitutive theories are quite often characterized in terms of the following bi-conditional: necessarily, *M*(*a*) iff *a* believes that *M*(*a*). The problem with this is that the simple truth of the bi-conditional does not explain the relation between *M*(*a*) and *a*'s higher-order belief that *M*(*a*). There is therefore no reason why such a bi-conditional could not turn out to be true of wholly distinct states, indeed no reason why it could not be accidentally true (for more on necessary connections holding between distinct states, see Wilson [2010]). This is why Wright's account does not need to deny the *Distinct Existence Thesis* — the heart of his proposal is that the bi-conditional is primitively true. In this formulation of the *Distinct Existence Thesis*, the qualification 'first-personally' is needed to rule out impertinent counterexamples generated by the referential opacity of the belief context: for example, a case in which *a* cannot remember who she is, but nevertheless believes of herself that she is in *M* (for discussion of these sorts of cases, see Rumfitt [1994]). For ease of prose, I will often leave the qualification implicit in what follows.

6. This is plausibly because the operator picks out (partially) the identity conditions of the entity in question. Therefore, in claiming that part of what it is to be X is to be Y, one is saying something stronger than that X is partially constituted by Y.

7. *Treatise* 1.4.5.5.

is, someone might think that, in ordinary circumstances, we are not ignorant of our own psychological states. Similarly, one might think that, in optimal conditions for self-reflection, we are not mistaken about our own psychological states (cf. Shoemaker 1990). One might therefore think that, because they are, in some sense, non-standard cases, the mere possibilities of ignorance or of error could be accommodated by a constitutive view that denies the *Distinct Existence Thesis*. We shall return to this issue in section 5.

Neither of these two considerations amounts to a decisive argument for the *Distinct Existence Thesis*. But I do think they suggest that it is plausible and intuitive. In that case, however, one might wonder why so many philosophers endorse constitutive theories. One motivation is the widespread conviction that first-person access is epistemically unique or distinctive. Some philosophers worry that if the *Distinct Existence Thesis* were true, then our first-personal way of knowing about our own psychological states would be based on causal relations but, in that case, it would no longer be epistemically distinctive. It would too closely resemble perceptual knowledge of the external world.

A second reason that one might endorse a constitutive theory is provided by a very influential argument presented by Sydney Shoemaker in a number of essays. Shoemaker's argument maintains that if a subject's psychological states were wholly distinct from her higher-order beliefs about them, then it would be possible for someone to suffer from a condition that he calls "self-blindness". A person is self-blind just in case she can know about her psychological states in *only* a third-personal way. Shoemaker argues that the possibility of self-blindness is absurd and that for this reason a subject's psychological states *cannot* be wholly distinct from her true higher-order beliefs about them. If his argument were sound, then the impossibility of self-blindness would show that some type of constitutive theory must be true.

The goal of this essay is to argue for a philosophical conception of self-knowledge that is consistent with the *Distinct Existence Thesis*, which is to say *not* a constitutive theory. In the first two sections, I shall argue that even if self-blindness is impossible the best explanation for this does not deny the *Distinct Existence Thesis*. In sections three and four, I draw on recent work on self-knowledge and rationality in order to present an account of first-person access that respects its epistemic distinctiveness, but is also consistent with the *Distinct Existence Thesis*. The account I present is inspired by Richard Moran's work on self-knowledge, and it maintains that a rational subject with the capacity to consciously self-ascribe a psychological attitude must ordinarily take her attitudes to depend on her assessment of reasons. I shall argue that meeting this condition on rationality requires one to have a capacity for first-person access which is epistemically distinctive in two respects. In the final section of the essay, I consider an objection to this proposal, and then I argue that, because the *Distinct Existence Thesis* allows us to make better sense of self-ignorance and error, we should prefer theories of self-knowledge that are consistent with it.

1. Self-Blindness

According to Shoemaker, "a self-blind creature would be one which has the conception of the various mental states, and can entertain the thought that it has this or that belief, desire, intention, etc., but which is unable to become aware of the truth of such a thought except in a third-person way" (1988, pg. 115). Self-blindness is supposed to be analogous to color-blindness. Just as a person who is color-blind can learn information about color in a non-standard way — through reading a book about colors, for instance — a person who is self-blind can learn about her own beliefs, desires, and intentions in a non-standard way, which is to say a completely third-personal way. Instead of having first-person access to her psychological states, the self-blind person will have to make observations of her behavior in order to know what she believes, wants, or intends. Moreover, self-blindness is "supposed to be perceptual or quasi-perceptual, rather than cognitive or conceptual" (1988, pg. 118). For this reason, self-blindness is a

condition of subjects who are at least as rational and conceptually sophisticated as any ordinary person.

Shoemaker actually tailors his self-blindness argument to different types of psychological states, but each variation rests on a version of a thought experiment that asks us to try to imagine a self-blind person with respect to a specific type of state.[8] In what follows, I will focus primarily on beliefs, but Shoemaker's arguments are similar for other attitudes. My use of the term 'psychological attitude' or 'attitude' is meant to apply only to attitudes that are sensitive to a subject's judgments, paradigmatically beliefs, desires, and intentions (cf. Scanlon 1998; Hieronymi 2005). These are the focus of Shoemaker's arguments, and they are also at the center of many recent discussions of self-knowledge (e.g., Bilgrami 2006; Boyle 2009, 2011; Fernandez 2013; Gertler forthcoming; Moran 2001, 2012; Parrott 2015). So, if there are any attitudes that are insensitive to judgment — perhaps implicit attitudes or what Gendler (2008) calls "aliefs" — these will fall outside the scope of this essay.

When he discusses belief, Shoemaker asks us to try to imagine a self-blind person that he names George. Shoemaker's first argument involves Moore's paradox. He suggests that, if George were self-blind, we should be able to imagine that "the total evidence available to a man at a given time should support the proposition that it is raining, while the total 'third-person' evidence available to him should support the proposition that he does not believe that it is raining" (1988, pg. 118). In this case, it would seem reasonable for George to assert a Moore-paradoxical proposition like 'It is raining, but I do not believe that it is raining'. But Shoemaker points out that any rational person can recognize a Moore-paradoxical assertion is inappropriate. Since we are supposing George to be fully rational, he will avoid asserting a Moore-paradoxical proposition. Hence, he would not attribute beliefs to himself in a manner that noticeably diverged from the way any of us would.

8. The different versions can all be found in Shoemaker (1994).

Shoemaker also presents what seems to be a second argument. He thinks that, because George has mastery of the concept of belief, the following two things will be true:

A) He will recognize that when asked "Do you believe that P?", he ought to answer "Yes" just in case he would answer "Yes" to the question 'Is it true P?'.
B) He will recognize the meaning of 'believe' and preface his assertions with 'I believe' in just the circumstances in which this is pragmatically appropriate.

Plausibly, A) follows from George's being a rational believer with the concept of belief. When asked whether or not you believe that *P*, like George, you usually consider the truth of *P*.[9] The concept of belief is that of an attitude responsive to the truth it represents, and understanding this connection between belief and truth is a large part of having that concept.

Having the concept of belief also means George can appreciate the relevance of pragmatic considerations for self-ascribing belief. For example, when I lose my keys, I have to look for them. Where are they? I don't really know, and there is no clear evidence indicating their location. But I have a hunch they are in my office. In this sort of case, my saying "I believe the keys are in my office" is a way of making a guarded assertion about the location of my keys. Since George has conceptual mastery, he could behave the same way I do when I lose my keys. Like me, he could say "I believe that the keys are in my office" in appropriate contexts.

On the basis of these arguments, Shoemaker concludes that George would self-ascribe beliefs in the exact same conditions any of us would. There would therefore "be nothing in his behavior, verbal or otherwise, that would give away the fact that he lacks self-acquaintance". When

9. This familiar point is found in discussions of self-knowledge that focus on its so-called "transparency" (e.g., Boyle 2011; Byrne 2005, 2011; Fernandez 2013; Moran 2001, 2012).

we try to imagine George, we can only imagine a person who reports beliefs just as we would, which means that we *cannot* really imagine a distinctively self-blind person. This, Shoemaker insists, just means that self-blindness is not a genuine possibility. If it were, "there should be something that would show, or at any rate provide good evidence, that someone was afflicted with such self-blindness" (1994, pg. 233).

One might worry that Shoemaker's argument is committed to some kind of behaviourism. But the claim he is making is not that George must have some type of first-person access because his observable behaviour indicates that he does. It may be that there are conditions in which it is in George's best interest to deceive others about what he believes, or cases where he simply doesn't care whether others know what he believes. Shoemaker's point is that, simply by virtue of being rational, George will be able to reason from his awareness of what he takes to be true, to various types of action that are rationally appropriate given what he takes to be true and, crucially, in a wide range of contexts this will include him self-ascribing beliefs. In Shoemaker's words, "he acts as if he believes that p, when he does so believe, which seems contrary to the supposition that he is self-blind with respect to his beliefs"[10] (2009, pg. 37).

It is important to keep in mind that George is supposed to be *completely* unable to become aware of his beliefs except in a third-personal way. Even if we can conceive of someone lacking first-person access to some of her beliefs, this is not self-blindness. We all lack that kind of access to some of our beliefs some of the time, but none of us are self-blind. To conceive of a self-blind person, we would need to imagine a person who is as sophisticated as we are psychologically, rationally, and conceptually, without any first-person access at all. Is this possible? I tend to agree with Shoemaker that it is not.[11] It is certainly difficult for me to imagine such a person. However, regardless of whether one could make a case for it, for the remainder of this essay, I shall stipulate that self-blindness is impossible in order to see what this means for the *Distinct Existence Thesis*.

10. For Shoemaker's own defense against the charge of behaviourism, see section 2 of his (2009).

11. Others disagree. See, for instance, Kind (2003) and Finkelstein (1999).

2. The Argument from Self-Blindness

Shoemaker presents the self-blindness argument against a model of self-knowledge that he labels the "broad perceptual model", which encompasses any theory of self-knowledge that is committed to the following two conditions:

> The "**causal condition**" that "our beliefs about our mental states are caused by those mental states" and
>
> The "**independence condition**" that "the existence of these states and events is independent of their being known in this way, and even of there existing the mechanisms that make such knowledge possible" (1994, pg. 271).

According to Shoemaker, if the broad perceptual model were true, then self-blindness would be possible. Yet, although this model is characterized by Shoemaker in terms of these two conditions, it is really just the "independence condition" that figures in his argument. According to Shoemaker, the "logical" possibility of self-blindness is a "consequence of the independence condition" (1994, pg. 273; cf. 1988). Thus, even though what he calls the causal condition plausibly entails the independence condition, it is worth keeping in mind that refuting Shoemaker's self-blindness argument would not demonstrate that the causal condition is correct. It would not demonstrate that a subject's psychological state *M* stands in some kind of *causal* relation to her higher-order belief that she is in *M*. Rather, it would only show that we do not yet have an argument for thinking that the two stand in a necessary relation.[12]

12. One view that therefore remains open is that *a*'s mental state *M* is a proper part of or partially constitutes her higher-order belief that she is in *M* (so long as the constitution relation here is not construed as identity). This would

Does Shoemaker's self-blindness argument refute the *Distinct Existence Thesis*? Shoemaker does not explicitly address the *Distinct Existence Thesis* as I have formulated it. Nonetheless, it might be thought that the *Distinct Existence Thesis* plausibly entails at least the **causal condition** that Shoemaker attributes to the broad perceptual model. This is because, as we have seen, the thesis commits one to thinking that one's being in a psychological state *M* and one's belief that one is in *M* are wholly distinct and therefore not necessarily connected. Since our first-personal way of knowing would therefore rest on contingent causal relations, the *Distinct Existence Thesis* would also plausibly entail the **independence condition.** It follows that if the *Distinct Existence Thesis* were true, self-blindness would be possible.

The self-blindness argument can be rendered in the form of *modus tollens*:

(1) If the *Distinct Existence Thesis* is true, then self-blindness is possible.
(2) Self-blindness is not possible.
Thus, the *Distinct Existence Thesis* is not true.

allow for the two entities to be construed as distinct in the sense required by Shoemaker's independence condition: *M* could exist independently from *a*'s higher-order belief that *a* is in *M* (and vice versa). This is the sort of contingent parthood relation that holds between my left thumb and me or between me and my family. It has no modal implications (even if it always holds in normal conditions). But one might naturally call it a "constitutive" theory. It is nevertheless quite different from the constitutive theories prevalent in philosophy which are clearly presented as having modal implications. Since I lack space to discuss this type of view in this essay, I shall briefly mention just one potential problem before setting it aside.

Many philosophers will want to hold at least some sort of weak supplementation principle on parthood relations. But I don't think it is clear what the other non-overlapping proper parts of *a*'s higher-order belief that *a* is in *M* could be (for some very intuitive considerations in support of weak supplementation, see Sider 2007). For instance, Shoemaker mentions things like rationality, intelligence, and conceptual capacities, but it seems to me these might potentially violate a standard anti-symmetry principle governing parthood — for instance, it is plausible that truly believing that one is in *M* is *part* of being rational.

Since we have already granted (2), this argument can be resisted only if premise (1) is wrong. Thus, we must inquire as to whether the *Distinct Existence Thesis* really does entail the possibility of self-blindness.

Shoemaker's argument for (1) proceeds primarily by analogy. Specifically, he appeals to perceptual knowledge to illustrate what it is for an entity to be independent from our way of knowing about it. It is clearly possible that material objects could exist without us and without any of our modes of perceptual access to them. This, Shoemaker claims, is a consequence of their being "logically independent" from our ways of knowing about them:

> The objects and states of affairs which the perception is of, and which it provides knowledge about, exist independently of the perceiving of them, and with certain exceptions, independently of there being things with the capacity for perceiving them or being aware of them. Thus trees, mountains, etc. can exist without there being creatures with the capacity to perceive them, and it is in principle possible for houses, automobiles and human bodies to exist in this way. [N.B. the exception is perceiving other perceivers.] (1994; pg. 254)

Just as trees and mountains could exist in a world without creatures capable of perceiving them, Shoemaker thinks that if beliefs were wholly distinct from our standard way of knowing about them, it would be possible for them to exist in creatures that lacked the capacity to access to them in a first-personal way. Thus, according to Shoemaker's argument, (1) is true because the *Distinct Existence Thesis* entails "that for each kind of mental fact to which we have introspective access, it is at least logically possible that there should be creatures in which such facts obtain, and who have the ability to conceive of them, but who are self-blind with respect to them" (1994; pg. 273).

This line of reasoning for (1) can be resisted. It is not true for any

entities, α and β, that if they are independent from each other, then there is a possible world in which α exists and β does not. If, for example, β were a necessary existent, then there would be no possible α-world that is not also a β-world, but not because the existence of α somehow depends on the existence of β.[13] Along similar lines, if two properties F and G are independent from each other, this means that *some* possible individual can exemplify F without G (and *vice versa*). But there may nevertheless be a certain kind of individual, or even a particular individual, that cannot exemplify F without exemplifying G. That will happen, for example, whenever G is an essential property of the kind or individual (though that is not the only possible way for this to be the case).

This is why Shoemaker's analogy with material objects is misleading. We tend to have the intuition that trees and mountains are substances, which suggests that they are independent not only from our capacities to perceive them, but from everything else as well. It is therefore quite easy to envision a world with trees and mountains but no "creatures with the capacity to perceive them". By contrast, psychological attitudes are not substances; they are properties of psychological subjects, some of which, like us, are rational subjects.

The previous section's reflections on self-blindness concern modality, not independence. If they are correct, they show that there is no possible world in which a rational subject lacks the capacity to access her beliefs in a distinctively first-personal way. But, to determine whether or not (1) is true, the relevant question is not *whether* this is the case, but *why* it is the case. One way to explain this necessity is by appealing to the nature of the underlying attitudes. Shoemaker typically writes as if he prefers this kind of explanation. For instance, he claims that a rational subject cannot be self-blind, because it is "the essence of many kinds of mental states to reveal themselves to introspection" (1994; 287). He also claims that denying the possibility of self-blindness amounts to holding a view about "the nature of certain mental states" (1988, pg. 31). However, rather than appealing to the nature of certain psychological attitudes, it might be that the capacity for first-person access is grounded instead in the nature of our rationality. That is, it may be that our rational nature, rather than the nature or essence of any psychological attitudes, is what explains why self-blindness is impossible.

To be fair, Shoemaker sometimes writes as if he has this last idea in mind. For instance, in one of his earlier papers, he says "it is of the essence of mind that each mind has a special access to its own contents, or more soberly expressed, that each person has a special access to his own mental states" (1988, pg. 115). He also makes several remarks about how what is "essential to a rational being" involves being "sensitive to the contents of one's belief-desire system in such a way as to enable its contents to be revised and updated in the light of new experience, and enable inconsistencies and incoherences in its content to be eliminated" (1994, pg. 285; see also 1990). In the following section, I shall draw on precisely this sort of idea to argue that the nature of rationality can fully explain the impossibility of self-blindness. I take my argument to be congenial to most of what Shoemaker says about rational agents and rationality. But, as I shall argue, if the impossibility of self-blindness can be fully explained in this manner, we do not have to accept (1). We are not committed to the premise that the *Distinct Existence Thesis* entails the possibility of self-blindness, if some other feature of rational subjectivity rules it out.

3. Rational Self-Awareness

Several philosophers writing on self-knowledge have recently stressed that a rational subject's psychological attitudes like belief, desire, and intention are normally responsive to reasons (e.g., Bilgrami 2006; Boyle 2011; Moran, 2001, 2012; Parrott, 2015). For example, Richard Moran writes that "I take what I believe to be answerable to my sense of reasons and justification, and I take myself to be responsible for making my belief conform to my sense of the reasons in favor or against" (2003, pg. 405). It should be fairly uncontroversial that, as a

13. For other examples and for discussion of the relation between modality and dependence, see Fine (1995).

rational subject, one's beliefs are usually sensitive to evidence and to justifying reasons that bear on the truths they represent. If I believe that *P* and am confronted with what I take to be a conclusive reason that *P* is false, I will, insofar as I am rational, immediately stop believing *P*. In this way, my assessment of the world, my take on reasons for or against the truth of *P*, makes an immediate and substantial difference to the existence and character of my belief. My appreciation of reasons for or against my beliefs directly affects them and has the potential to change them. In this section, I would like to suggest that this explains why a rational subject with the ability to consciously self-ascribe beliefs must have a capacity for first-person access to them.

Tyler Burge has stressed that having this mode of epistemic access is necessary for an individual to engage in what he calls "critical reasoning" with respect to her beliefs.[14] Burge argues that "it is constitutive of critical reasoning that if the reasons or assumptions being reviewed are justifiably found wanting by the reviewer, it *rationally follows immediately* that there is prima facie reason for changing or supplementing them, where this reason applies within the point of view of the reviewed material (not just within the reviewing perspective)" (1996, pg. 109; cf. Burge 1998 and Gertler forthcoming). Someone reasoning critically in Burge's sense must focus her attention on her beliefs so that she can rationally evaluate them. This will sometimes culminate in a judgment to the effect that a particular belief is warranted or not by reasons, a judgment that Burge notes necessarily has the potential to immediately change the original state. Burge thinks that if our way of knowing about our own beliefs always rested on behavioral evidence, "there would never be an immediate rationally necessary connection" (1996, pg. 109) between what we deliberatively judge we ought to believe and what we in fact believe. Why not?

In order for someone to engage in critical reflection, it is clear that she must be aware of her beliefs. Suppose she is indirectly aware of them in a more third-personal manner. When someone takes up this sort of perspective, she treats her own beliefs similarly to the way she does those of others, which is to say that she treats them like facts that are not necessarily determined by her assessment of what is most reasonable to believe (cf. Moran 2012). This is because, from a third-person point of view, a person's best judgments about what her beliefs ought to be does not *settle* what they in fact are, and this is why critical judgments made from a third-personal standpoint lack an "immediate rationally necessary connection" to the subject's beliefs. They are judgments made from a perspective that treats one's beliefs as things that might not be determined by one's rational judgments about what she ought to believe.

In fact, for this reason, it seems to me that Burge overemphasizes the importance of critical reflection.[15] It rather seems to me that a rational subject who has the capacity to consciously self-ascribe beliefs requires first-person access to them, whether or not she ever subjects them to deliberative or critical evaluation. This is because, from a rational subject's point of view, one's beliefs depend on her having adequate reasons for them.[16] This is fundamental to the nature of the first-person perspective of a rational believer. From that

14. Shoemaker gives a similar argument in his 1990, 1994, and 2009.

15. Cf. Moran (2012). In contrast to Burge, Moran claims that "the non-observational character of self-knowledge with respect to actions and attitudes is tied to their being expressions of the rational, active side of one's nature" (2012, pg. 220; cf. 2001). He goes on to explicate this in terms of what he calls the "transparency condition".

16. What about groundless convictions, like my belief that the Lions will win the Super Bowl? These are surely not formed deliberatively on the basis of reasons or evidence. Even so, for a rational subject, we might think that these sorts of convictions depend on her having adequate reasons in the sense that once she becomes aware of compelling evidence that the belief is false (e.g., the Lions fail to make the playoffs), she will stop believing that the Lions will win the Super Bowl. It is worth noting that we might also think such a subject does not have the same kind of epistemic access to her groundless convictions. For example, it seems less likely that she could know that she believes that the Lions will win the Super Bowl on the basis of what Moran (2001) calls the "transparency method". For further discussion of groundless convictions, see Martin (1998).

perspective, if I take myself to believe that *P* — which I definitely do if I consciously attribute the belief to myself — it is *only* because I take there to be some good reason or set of reasons in favor of the truth of *P*. Note that this does not mean that my belief that *P* must be the result of any conscious, deliberative, or reflective process; as Burge rightly notes, "much of our reasoning is blind, poorly accessible, and unaware" (1996, pg. 99; cf. Kornblith 2012). Nevertheless, as a rational subject, when I *attribute* a belief to myself, I am attributing an attitude that I at least tacitly conceptualize as being appropriately grounded in reasons for believing, even in cases where I cannot articulate what those reasons are.

Some philosophers think this conception of rational belief is too demanding. For instance, Quassim Cassam claims that "much of the time, our reasoning isn't guided by an appreciation, use, and assessment of reasons and reasoning as such. We are reluctant to criticize our reasons, and we do not guard against possible sources of bias" (2014, pg. 17). So it is worth emphasizing that the basic idea that a rational subject takes her beliefs to depend on adequate reasons for believing is fairly minimal. Insofar as I am rational, when I self-ascribe the belief that *P*, I do not need to grasp precisely which considerations are my reasons or basis for believing that *P* (I may have forgotten them), nor do I have to have formed the belief on the basis of conscious deliberation about some set of reasons or evidence (I often don't do this). Rather, as a rational believer, I am simply committed to there being *some* sufficient reason(s) in favor of believing that *P*, and this is because my rational beliefs must be sensitive to any judgments about reasons that I *would* make.[17] So, for instance, if I were to judge that there were absolutely no good reasons to believe that *P* (perhaps I have just learned that the source of my information is not trustworthy), then, insofar as I am rational, I will immediately stop believing that *P* (cf. Parrott, 2015). As Burge suggests, there must be a rationally *immediate* connection between this sort of judgment about reasons and what I actually believe. My point is that this connection must hold in cases other than those where a subject critically reflects upon her beliefs.

If a rational subject's ability to self-ascribe beliefs involves conceiving of them as attitudes that depend exclusively on reasons, then this plausibly requires one to have a mode of epistemic access to them that is not based on behavioral evidence. If my way of self-attributing beliefs rested entirely on third-personal ways of knowing, then it would mean that, from my own perspective, my belief that *P* might depend on something other than what I regard as adequate reasons for holding it. This is because behavioral evidence in favor of the proposition that I *believe* that *P* is typically not evidence for the truth of the proposition *P*. So, if I base my attribution on such behavioral evidence, on something other than a reason in favor of *P*, it amounts to admitting that my belief might not depend exclusively on reasons for believing *P*. In that case, my act of self-attribution would leave open the possibility that there are not good reasons for believing that *P*.

An example may help make this point clearer. Suppose that I come to know that I believe that my neighborhood is unsafe on the basis of behavioral evidence. I recognize, for instance, that I check the locks on my windows and doors much more frequently than in any other place I have lived. I also notice that I walk extremely quickly through my neighborhood and regularly glance over my shoulder. This kind of behavior is good evidence that I believe my neighborhood is unsafe and so good evidence for attributing this belief to myself. But it is not good evidence that the neighborhood is actually unsafe. Indeed, the behavioral evidence for what I believe is consistent with there being no reasons at all to think my neighborhood is actually unsafe. By self-attributing a belief in this way, I leave open the possibility that what I believe is not determined by what I think I ought to believe. My point is that when a person relates to her own beliefs in a first-personal way,

17. Cassam (2014) emphasizes several ways in which human beings are sometimes less than rational, for instance by having recalcitrant beliefs which persevere in the face of counterevidence (cf. Bortolotti 2010). These are interesting cases, but are beyond the scope of this essay. Again, the question is whether, insofar as one is a rational agent, one must have a capacity for first-person access. If so, this is compatible with one lacking first-person access to a range of attitudes.

she must take this *possibility* to be closed off. From that perspective, she must take herself to believe that her neighborhood is unsafe only because that belief is adequately supported by reasons and that requires her to have a capacity for epistemic access to her beliefs that is not based on publically available behavioral evidence.[18]

But couldn't I have some reasons, perhaps even excellent ones, for thinking that the beliefs I ascribe even from a third-person perspective are based on good reasons? For instance, especially since I don't remember the basis for much of what I believe, I might reasonably just take myself to have a general reason for thinking that all my beliefs are based on good reasons. So, even if I were to self-ascribe the belief that *P* from a third-person perspective, I would nonetheless take that belief to be appropriately grounded in reasons for believing *P*. It seems unlikely that the first-person method of self-ascription would give me any better reasons for thinking that my belief is appropriately grounded in reasons.

The point, however, is not about whether one has *epistemic* reasons for thinking that one's beliefs are actually based on good reasons. It may be true that I have reasons for thinking that everything I believe is based on good reasons for believing. But the point is about the necessary "rational immediacy" of the connection between a rational subject's judgments about reasons for believing and the things she actually does believe. It is internal to the first-person perspective of a rational subject that the beliefs one self-ascribes are immediately sensitive to the judgments one makes.

By contrast, self-ascribing the belief that *P* on the basis of behavioural evidence leaves open the possibility that what I believe diverges from my own rational assessment of reasons right now, at the moment of self-ascription. A third-personal mode of self-ascription leaves open the possibility that, if I were to reconsider the question of whether *P*, I might come to a conclusion that *diverges* from what I actually believe. That is a possibility that must be closed off in order, as Burge says, for rational judgments or evaluations to *necessarily* have an immediate consequence on what I believe. The difference between the third person and the first person is therefore not one of acquiring better epistemic reasons for taking one's beliefs to be based on sufficient reasons or evidence. It is rather that only from the first-person perspective do one's rational judgments immediately affect what one believes.

In response, someone might object that just as considering one's total evidence is the most rational way to form beliefs about one's neighborhood, it is also the most rational way to form beliefs about one's beliefs. Indeed, it could even be argued that if a person is attempting to learn about what she *already* believes as opposed to what she ought to believe, she should consider every bit of available evidence so as not to accidentally change what she believes (Shah and Velleman 2005). I think there are cases where it is responsible to base one's self-ascriptions on behavioral evidence. Prior to noticing my nervous habits, I was not aware of my belief that my neighborhood is unsafe. Nevertheless, that belief played a pretty important role in my life. It may be that the only way I could ever have learned about it was by noticing my nervous reactions. If so, it seems responsible to acquire knowledge of this belief on that basis. However, it is important to realize that this point does not generalize. It is not that a rational agent cannot sometimes self-ascribe beliefs in a third-personal way. But she can do so only by distancing herself or disengaging from the more fundamental first-person standpoint of a rational agent (cf. Moran 2012). If she *always* acquired knowledge of her beliefs in a third-personal way, they would remain at too far a distance from her sense of reasons for holding them.

The proposal sketched in this section is one way of explicating the idea that the impossibility of self-blindness can be accounted for by the nature of rationality. But absolutely nothing about it implies that a subject's psychological states must be related in a way that is incompatible with the *Distinct Existence Thesis*. If our rational subjectivity requires first-person access to our own attitudes, then

18. Moran (2001), I think, is getting at a similar thought when he describes first-personal self-knowledge as a "rational requirement".

we do not have a reason for denying that those attitudes are wholly distinct existences.

4. Dependence

The previous explanation of first-person access rests on the claim that a rational subject's beliefs depend on her sense of reasons for them. It is important to clarify this. A number of philosophers employ the phrase 'sense of reasons', but in doing so, they can mean very different things.[19] One might mean to refer to a subject's actual judgments, either implicit or explicit, about reasons for or against her beliefs. Or, alternatively, one might mean to refer to a subject's dispositions to judge certain things about reasons. I prefer the second interpretation of 'sense of reasons'. Notice, however, that even if we were to adopt the former, the claim that a subject's beliefs depend on her actual judgments about reasons is not obviously a claim with any modal implications. If an agent explicitly and consciously judges that there are conclusive reasons against the truth of *P*, her belief that *P* might cease to exist, but it might also persist despite her well-considered judgment (cf. Cassam 2014, chapter 2). This sort of thing should be familiar, and it shows, I think, that a rational subject's beliefs can persist independently of her actual judgments about reasons.[20]

On my preferred interpretation, the central claim of the previous section is that a rational subject's belief that *P* depends on her being disposed to judge in the right circumstances that reasons adequately support the truth of *P*. Although this is consistent with the *Distinct Existence Thesis* as formulated, one might naturally worry that it goes against the spirit of the thesis. Specifically, one might take the proposal in the previous section to be making a claim about a different sort of necessary constitutive relation.

To see this, we can express the proposal from the previous section using the 'part of what it is to be' operator:

> Belief Rationality 1: *Part of what it is for a* to believe that *P* is for *a* to be disposed to make the appropriate judgments about reasons.

Indeed, since beliefs are frequently thought to be dispositions or sets of dispositions, it is quite easy to see how they might be either identical to, or partially constituted by, a subject's dispositions to judge certain things (cf. Schwitzgebel 2002).[21] Moreover, if we interpret the previous section's proposal along these lines, Belief Rationality 1 would have modal implications; for instance, necessarily (if *a* believes that *P*, then *a* is disposed to judge [in the right circumstances] that *P* is true).

I don't think it is obviously wrong to think that part of what it is for a rational subject to believe that *P* is for her to have dispositions to judge certain things. Nevertheless, I think this is not the best way to understand the relation between belief and judgment. The issue comes down to whether or not we are inclined to think that a rational subject can believe that *P* without having the appropriate dispositions. For instance, could a rational subject believe that *P* and be disposed to judge that evidence conclusively shows *P* is true but then, for various reasons, lose this disposition while nevertheless retaining the belief? If this sort of thing can happen, then it does not seem quite right to think that *part of what it is* for *a* to believe that *P* is for *a* to be disposed to judge that the evidence shows *P* is true.

One might object that this sort of behavior would indicate some kind of irrationality on the part of the subject. That is, someone who

19. This language is prominent in Moran (2001, 2012). However, it is not clear to me which of the interpretations presented in this section, if either, Moran would favor.

20. It is worth noting that even if one were inclined for some reason to think a subject's beliefs did depend on her actual judgments, this would still be very different from the sort of relation we have been considering in this essay insofar as it would be compatible with the subject's beliefs being wholly distinct from her higher-order beliefs about them.

21. One might also think that beliefs are constitutively related to other beliefs such that, for instance, part of what it is for *a* to believe that *P* is for *a* to believe that *Q*. It seems to me that the reasons given in this section against Belief Rationality 1 would also be reasons to avoid this sort of view in favour of one analogous to Belief Rationality 2.

forms a belief on the basis of a deliberative judgment that *P* is true, but then retains that belief despite losing her disposition to judge that *P* is true, might seem to be manifestly irrational. But since *a* is by hypothesis a rational subject, it does not matter whether an *irrational* believer could continue to believe something without retaining the sorts of dispositions one finds in rational subjects. Yet, it does not seem to me that losing the disposition to judge that *P* is true, or is adequately supported by evidence, automatically impugns the rationality of a believer. Indeed, in paradigmatic cases of rational belief revision, there will be an interval where a subject both believes that *P* and lacks the dispositions to judge that *P* is true, or likely to be true, or adequately supported by evidence. It seems to me that sometimes that interval can be quite significant.

Let's return to the belief that my neighborhood is unsafe. Suppose now that I believe this not out of fear or paranoia but because it is actually not safe. Imagine that the rate of violent crime is unusually high, that burglaries are common, and that most of the neighborhood residents are armed. In such a scenario, it is reasonable to think I would both be disposed to judge and rationally believe that my neighborhood is not safe. Let's suppose this is true. But now imagine that over a number of years my neighborhood is the target of major government intervention. Police presence increases and there is an influx of private investment. Both unemployment and criminal activity decline and, over time, the neighborhood slowly becomes safe. Noticing these gradual changes, I get to the point where I am no longer disposed to judge that my neighborhood is unsafe. For example, if someone were to ask me whether it was, I would say "No." Despite my acknowledgment of the newfound safety of the neighborhood, it seems possible that I continue to *believe* that the neighborhood is unsafe. This might involve me manifesting behavior such as glancing nervously over my shoulder or installing extra locks on my house. We know that an individual can have beliefs like this, beliefs that are discordant with what they are disposed to judge (cf. Peacocke 1998; Cassam 2014). It also seems to me that beliefs of this sort can persist for a fair amount of time. Indeed, as Cassam notes, "the more long-standing and deeply embedded your belief that P the harder you may find to shake it off when confronted by evidence which you realize undermines it" (2014, pg. 23). This suggests that one's belief can persist even after one's dispositions have changed.[22]

If, as a rational agent, I am able to retain the belief that my neighborhood is unsafe even when I am no longer disposed to judge that it is, then it is less plausible to think that the belief itself depends on my dispositions to judge. Of course, once I lose the appropriate dispositions, my belief plausibly becomes irrational, even by my own lights, which means it changes in an extremely significant respect. But it notably does not cease to exist. This suggests that it is the rationality of my belief, rather than its existence, that depends on my dispositions to make the appropriate judgments. I therefore propose that we understand the idea that a rational subject's beliefs depend on

22. Someone might argue that, even in cases like my neighborhood example, a subject does not really lose the dispositions to judge that *P* is true or supported by good reasons; it is just that the dispositions are masked in some way. One might therefore argue that even after the neighborhood becomes safe, I retain the disposition to judge it is unsafe; it is just that that disposition is inhibited. Although this alternative way of describing the case is available, I think there are at least two reasons to resist it. First, we might think an individual simply cannot have directly opposing dispositions. For example, a person cannot simultaneously be disposed to blink and be disposed to not blink in the same circumstances (cf. Handfield and Bird, 2008). Secondly, with respect to an ability to act in a certain way, one might be disposed to exercise the ability in a specific manner or toward a particular end (e.g., one might be disposed to run slowly, or for ten miles); however, in such cases, we tend to think that the individual's disposition can be masked only by something external to the agent, not by one of the agent's intrinsic properties. We can perhaps imagine a case in which someone is disposed to judge that *P* but bizarre events bring about the opposite judgment (this is the sort of thing that happens in Frankfurt-style cases). Yet, if nothing extrinsic to the agent causes her to judge that *P*, if that judgment is her successful intentional action, it is hard to see how the person could also have some masked disposition to judge that not *P*. Similarly, if a person deliberates on the safety of the neighborhood and judges it to be safe, there is something odd about thinking they have somehow failed to exercise the disposition to judge it unsafe.

her sense of reasons as claiming that one's beliefs are rational in virtue of one's dispositions to make appropriate judgments about reasons.[23]

We can still use the 'part of what it is to be' operator to express this sort of dependence:

> Belief Rationality 2: Part of what it is for *a* to rationally believe that *P* is for *a* to be disposed to make the appropriate judgments about reasons.

Like its ancestor, Belief Rationality 2 will have modal implications. For instance, necessarily (if *a* rationally believes that *P*, then *a* is disposed to judge [in the right circumstances] that *P* is true). The crucial difference between this formulation and Belief Rationality 1 is that 'rationality' now serves to modify *a*'s belief that *P*. This reflects the fact that it is the *rationality* of one's belief that depends on one having the appropriate dispositions to judge that *P* is true, or likely to be true, or supported by evidence. If one loses those dispositions, then one's belief may persist, but it would be irrational. However, an instance of an irrational belief does not indicate that the *subject* of that belief fails to be rational. Rational subjects sometimes believe irrational things — indeed they sometimes *consciously* believe irrational things. Our rationality is far from perfect.

My thinking about dependence in this manner is motivated, in part, by the thought that rationality does not demarcate a natural psychological kind.[24] So, although my rationally believing that *P* and my rationally desiring that *P* are *ways* in which I exemplify a psychological property (i.e., desire or belief), rational belief and rational desire are not themselves psychological kinds. If this is right, it means that a subject's way of relating to her attitudes determines whether or not they are rational, not the underlying natures or essences of those attitudes.[25] It would therefore be wrong to conclude from the fact that a rational agent must have certain kinds of dispositions that this is because the existences of certain kinds of attitudes in her psychological life are, by their very nature, necessarily connected to those dispositions.

We are now in a position to see a second respect in which first-person access is epistemically distinctive when compared to the kind of epistemic access we have toward the psychological attitudes of others (the first, which we have already seen, is that it is not based on behavioral evidence). It is plausible that if someone has a capacity for first-person access, which is an epistemic capacity, she will thereby be disposed to form higher-order beliefs about her own psychological attitudes. Why is that? Generally, it seems that whenever a capacity does not require exercising volition, having the capacity to Φ *entails* being disposed to Φ.[26] Moreover, since exercising one's epistemic capacity for first-person access would involve forming a higher-order belief about one's attitudes, it means that anyone with a capacity for first-person access would be disposed to form higher-order beliefs about her own psychological attitudes. However, I have also argued that it is *necessary* that a rational subject have first-person access to her psychological attitudes, which suggests the following line of argument:

23. Some people (but only some) use the term 'grounding' for the type of relation I have in mind. For further discussion, see Audi (2012) and Rosen (2010).

24. This may mean that my conception of rational attitudes is committed to what Boyle calls an "additive theory" of rationality (2016). Boyle raises two objections to additive theories, which I think could be extended to the sort of view I'm sketching in this essay. But I lack space to address them in this essay.

25. But what if a subject's judgments are irrational? Wouldn't making an irrational judgment about reasons determine that a subject's beliefs are also irrational? I think it depends. A subject's belief can be epistemically irrational if it is formed on the basis of an epistemically irrational judgment, which is not correctable because, for example, we cannot get the subject to properly evaluate the available evidence (there is a question as to when this sort of recalcitrance becomes delusional; cf. Bortolotti 2010). This kind of irrationality of one's belief is derived from the epistemic irrationality of one's judgment. Part of what the discussion in this essay brings out is that there is a different kind of irrationality implicated in cases where one's beliefs come apart from what one is disposed to judge. This sense of irrationality might best be characterized in terms of a kind of dissociation, or alienation, but it is clearly possible even in cases where a subject's judgments about reasons for belief are epistemically flawless.

26. This assumption that first-person access is a non-voluntary epistemic capacity is one reason to think that what Shoemaker calls "first-person agnosticism" (1990) is impossible.

(3) Necessarily, for any rational subject *a* and psychological attitude *M*, if *M*(*a*), then *a* has a capacity for first-person access to *M*(*a*).

(4) Necessarily, for any rational subject *a* and psychological attitude *M*, if *M*(*a*) and *a* has the capacity for first-person access to *M*(*a*), then *a* is disposed to believe (first-personally) that *M*(*a*).[27]

First-Personal Dispositions: Necessarily, for any rational subject *a* and psychological attitude *M*, (if *M*(*a*), then *a* is disposed to believe [first-personally] that *M*(*a*)).[28]

Earlier I conceded to Shoemaker that it is impossible for a rational subject to lack the capacity for first-person access, but it now seems that this would mean it is impossible for a rational subject to lack a disposition to form true higher-order beliefs about her psychological attitudes. Because the first-person mode of epistemic access is something that a rational subject *must* have toward her attitudes, and because it implies the existence of this disposition, it looks different from the types of epistemic access we have to the attitudes of others. The epistemic capacities that allow me to access your attitudes and form beliefs about them do not seem to involve me having any disposition to form true beliefs about the attitudes you have — certainly not as a matter of necessity. Therefore, since the view I sketched in the previous section implies *First-Personal Dispositions*, it can be thought to validate the intuition that first-person access to one's own mental states is epistemically distinctive — it is a necessary epistemic capacity for rational subjects that brings with it a disposition to form true beliefs about one's own psychological attitudes. Although having this

27. There will naturally be instances in which this disposition is masked. I am grateful to Nick Jones for his assistance in helping me formulate this argument.

28. It is worth keeping in mind that *First-Personal Dispositions* applies to judgment-sensitive attitudes. What should we think about a rational agent's sharp credence like .4532? Is this a judgment-sensitive attitude? This is a good question, and it will be addressed in the following section.

capacity does not itself imply anything about the epistemic status of one's higher-order beliefs, and in particular it does not imply that they are epistemically authoritative or privileged, it is nevertheless a peculiar epistemic capacity insofar as a rational agent can exercise it only with respect to her own attitudes. Nevertheless, even if one denies that this feature of *First-Personal Dispositions* is sufficient to capture the special epistemic status of the first person, what is crucial for the purposes of this essay is that the conception of first-person access embodied in *First-Personal Dispositions* is compatible with the *Distinct Existence Thesis*.[29]

5. Luminosity and Fallibility

In the previous section, we saw that a rational subject with a capacity for first-person access will be disposed to form higher-order beliefs about the psychological attitudes she actually has (*First-Personal Dispositions*), which means that those higher-order beliefs will be true. Some philosophers might worry that this conclusion would mean that a certain range of psychological attitudes are self-intimating, since it claims that, as a matter of necessity, if a rational subject exemplifies one of these attitudes, she will be disposed to believe (truly) that she

29. It is worth noting that, according to this view, it is much less clear that first-person access is epistemically distinctive when we compare it to perception. To the extent that perception is an epistemic capacity, having a capacity for perceptual access to some range of facts plausibly entails being disposed to form beliefs about them. We could then construct a parallel argument for creatures that we define as "perceptual subjects", subjects whom we define as *necessarily* having a capacity for perceptual access. If perceptual access is also epistemically "direct", or not based on evidence or inference, which seems extremely plausible, then it would be epistemically analogous to first-person access. Thus, if one thinks first-person access must be epistemically different from perceptual access, one will think *First-Personal Dispositions* does not secure enough epistemic distinctiveness. Often, Shoemaker's self-blindness argument is cited in opposition to views that assimilate first-person access to perception, but, as we saw earlier, that argument does not support the conclusion that such assimilation makes self-blindness possible. Are there other reasons to sharply distinguish the epistemic properties of first-person access from those of perception? Perhaps, but addressing this question is beyond the scope of this essay.

does. In other words, some might worry that *First-Personal Dispositions* would make certain attitudes luminous, at least for rational subjects. As Williamson defines it, an attitude is luminous "if and only if, necessarily, whenever it obtains, one is in a position to know that it obtains" (2008, pg. 279). If one is convinced by Williamson's influential argument against luminosity, then one will want to insist that no psychological attitude is luminous.[30] This is an important objection, because, as we shall see, it is partially with respect to the possibility of luminosity that the sort of account presented in this essay diverges from mainstream constitutive theories of self-knowledge.

The reason someone might suspect that *First-Personal Dispositions* implies that a rational subject's attitudes are luminous is because she thinks that being disposed to believe that one exemplifies an attitude that one actually has, as *First-Personal Dispositions* claims, is equivalent to, or at least implies, being in a position to know that one exemplifies that attitude. That is, one might reason as follows:

(5) Necessarily, for any rational subject a and psychological attitude M, (if $M(a)$, then a is disposed to believe [first-personally] that $M(a)$). (*First-Personal Dispositions*)

(6) Necessarily, if, whenever $M(a)$, a is disposed to believe that $M(a)$, then a is in a position to know that $M(a)$.

(7) Therefore, necessarily, for any rational subject a and psychological attitude M, (if $M(a)$, then a is in a position to know that $M(a)$).

This argument would establish that the psychological attitudes of rational subjects are luminous (7).

In order for this argument to be credible, we would at least need to assume that a's disposition to believe that $M(a)$ is sufficiently reliable. *First-Personal Dispositions* claims that a rational agent will be disposed to believe that she exemplifies a psychological attitude whenever she does. But it does not say anything about dispositions one might have to form beliefs in other conditions. So, nothing in the argument excludes a from being disposed to form the higher-order belief that $M(a)$ in cases where it would be false. However, if that were the case, (6) would be quite obviously false. Thus, one would want to insist that a's disposition to form the belief that $M(a)$ whenever a is in M is epistemically reliable. Alternatively, one could argue that a rational subject is disposed to form accurate higher-order beliefs about *absences* of psychological attitudes, which would have the consequence of ensuring reliability as long as we ban explicit contradictions.

However, even if we were to make one of these two assumptions, (6) would still be false. Being disposed to truly believe that one exemplifies some attitude is not sufficient for being in a position to know that one does. This is because we are sometimes disposed to believe that a certain condition obtains in cases where we would have the same disposition in very similar cases where the condition does not obtain. In the good case, where a is disposed to believe that $M(a)$ and $M(a)$ is the case, her belief will be true, but the truth of this belief would be accidental if she were to have the same disposition in nearby cases where it is not the case that $M(a)$. Therefore, the disposition to form a true belief about one's psychological state in the good case would not be sufficient to put one in a position to know in that case. Indeed, this is one of the central lessons of Williamson's argument against luminosity.

Consider a series of times $t1, t2, \ldots tn$, such that one rationally believes that P at $t1$ and no longer believes it at tn. Suppose further that one's degree of confidence in P gradually diminishes over the course of this interval. There will then be some time ti such that one's one's confidence is still high enough to count as believing that P at ti but then, at $ti{+}1$, one's degree of confidence drops below the threshold for believing that P. Assuming that *First-Personal Dispositions* is true,

30. Williamson's (2000) version of the argument is presented in terms of sensations. Offhand, one might therefore think it applies only to phenomenal states. However, as we will see, the argument can be extended to attitudes like belief (cf. Silins 2012).

we can stipulate that at *ti* one is disposed to believe that one believes that *P*, but, as Williamson argues, one's higher-order belief at *ti* would not constitute knowledge, because it could easily have been false. It is extremely likely that one would retain the disposition to believe that one believes that *P* at *ti+1*, since the difference between one's degrees of confidence at *ti* and *ti+1* is likely to be indistinguishable. But if one would retain this disposition at *ti+1* (where, by hypothesis, the relevant condition does not obtain), then one would not be in a position to know that the relevant condition obtains in *ti*. So the disposition to believe that one has a psychological attitude whenever one does have it is not sufficient to put one in a position to know that one does. *First-Personal Dispositions* does not entail that any type of psychological attitude is luminous.

This formulation of the anti-luminosity argument raises two interesting questions about the status of *First-Personal Dispositions* that are worth further consideration. First, the subject in the example is not sensitive to fine-grained changes in her degrees of confidence between *ti* and *ti* + 1. This raises a question about whether degrees of confidence are even the sort of attitude that falls within the proprietary domain of first-person access. If they are not, then it would not indicate anything about one's capacity for first-person access if one did not have accurate beliefs about fine-grained credence in the truth of *P*. There are certain properties of material objects that cannot be perceptually accessed, but that doesn't mean our visual system is impaired.[31]

Second, according to the argument, in the borderline case *ti+1*, the subject has a psychological attitude that is by hypothesis not a state of believing that *P*. We might think of it as a state of suspended judgment, or agnosticism about *P*. But this means that *prima facie* the borderline case might look to be inconsistent with *First-Personal Dispositions*. The latter would imply that at *ti+1*the subject would *have* a disposition to believe that she is agnostic about *P*. However, according to the anti-luminosity argument, at *ti + 1* the subject retains the disposition from *ti*, which is to say that she is disposed to believe that she *believes* that *P*. This might be taken to indicate that there is something wrong with *First-Personal Dispositions*.

If we want to maintain *First-Personal Dispositions*, I think we should admit that, in a borderline case like *ti + 1*, a subject can have two dispositions, specifically the disposition to believe that she believes that *P* and the disposition to believe that she is agnostic about *P*. Having these two dispositions is not contradictory, so it is not impossible for a subject to have them both. There may seem to be something slightly odd about having both of these dispositions simultaneously, but let's keep in mind that borderline cases are non-standard.

To help alleviate the sense of oddity, let me clarify my conception of first-person access. On the sort of view presented in this essay, a rational subject must have a capacity for a special mode of epistemic access to her psychological attitudes. In the previous section, I argued that part of having this capacity involves having a general disposition to form higher-order beliefs about one's attitudes. In the majority of cases where this disposition is exercised, a subject will actually form a higher-order belief about the attitudes she actually has. So, in ordinary circumstances, her higher-order beliefs will be true. However, this disposition is not always exercised, and it can also be masked. This means that there may be cases where a rational subject will have the disposition to form a higher-order belief about her psychological attitude but not exercise it by actually *forming* a belief about the attitude she has. I would suggest that this is what happens in the sorts of borderline cases we find in Williamson's anti-luminosity argument. If I am in a borderline case where it is very difficult to distinguish whether or not I believe that *P* or I am agnostic about *P*, it does not seem that strange to think I may have both a disposition to believe that I believe that *P* and a disposition to believe that I am agnostic about *P*. From

31. If we think of credences as distinct kinds of psychological states, it may be sufficient to secure rationality that a subject is disposed to form true beliefs about intervals of probability. This would require us to modify *First-Personal Dispositions* somewhat, but it would be in the same spirit. But it may also be better to think of a subject's degrees of confidence not as credal states but as beliefs about probabilities; see for instance the arguments of Friedman (2013) and Sturgeon (2015). Thanks to an anonymous reader for raising the issues discussed in this paragraph.

this, nothing would follow about what a rational subject is disposed to believe in non-borderline cases. It therefore remains plausible that rational subjects are typically in a position to know about their own attitudes. The point of the anti-luminosity argument is simply to show that no psychological attitude *guarantees* that one will be in such a position.

By contrast, many constitutive theories do tend to maintain that at least some psychological attitudes are luminous (e.g., Bilgrami 2006; Coliva 2009, 2012; Shoemaker 2012; Zimmerman 2006). This should not be surprising. Taking one's psychological attitudes to be necessarily connected to one's higher-order beliefs certainly makes luminosity seem attractive.[32] For instance, if, following Boyle, we were to think that one's attitudes are simply *identical* to one's true higher-order beliefs about them, it would be very hard to see how one could fail to be in a position to know about an attitude whenever it obtained. By contrast, because it does not posit any necessary relation between a subject's psychological attitudes and her higher-order beliefs about them, the framework presented in this essay can simply deny luminosity. But is this a reason to prefer it?

Instances of self-ignorance present one reason to be suspicious of a constitutive theory committed to some form of luminosity. There is plenty of evidence demonstrating that individuals often have psychological attitudes that they are completely unaware of having (for a survey of some relevant experimental work, see Jost et al. 2009). To modify an example of Peacocke's (1998), I might have a belief that degrees from my own university are far superior to those from other institutions. Nevertheless, were someone to ask me what I believe, even if I were to consider the matter carefully, I could fail to be aware of this belief. The possibility of ignorance of one's own attitudes suggests that they are not luminous.

The luminosity of psychological attitudes is incompatible with the first clause of the *Distinct Existence Thesis*, the clause which denies that part of what it is to be in *M* is to believe (first-personally) that one is in *M*. So, it is possible for a constitutive theorist to reject luminosity but deny the second clause of the *Distinct Existence Thesis*. This sort of constitutive view would be committed to the claim that at least some of a subject's higher-order beliefs are infallible, because it would hold that *part of what it is to be* such a higher-order belief is for the subject to exemplify the embedded attitude. However, it would not be committed to the luminosity of any first-order attitudes and so could easily accommodate the possibility of self-ignorance

However, there is a parallel reason to be suspicious of a constitutive theory committed to some sort of infallibility, namely that it seems that we can be mistaken about our own psychological attitudes (cf. Snowdon 2012). This suggests that our first-personal way of knowing, however special it may be, is fallible. Notably our commonsense picture of positive self-deception is that it consists in a subject incorrectly believing that she has a particular attitude that she lacks.[33]

Constitutive theorists are naturally aware of these putative counterexamples, and they seek to accommodate them in one of two ways: First, several constitutive theorists restrict the scopes of their theses to one or more privileged kinds of attitude. For instance, Bilgrami and Coliva both distinguish a special kind of attitude that they call "commitment" (which includes ordinary beliefs and desires) and argue that there is a necessary constitutive connection between a subject's psychological attitudes and her higher-order beliefs only when the former are "commitments" (Bilgrami 2006; cf. Coliva 2012). According to this proposal, cases of self-ignorance involve an entirely different kind of state, a kind that, as Bilgrami says, "cannot possibly

32. For a nice argument that constitutive views are nevertheless susceptible to Williamson's anti-luminosity argument, see Srinivasan (2015). The most popular constitutive theories discussed in the literature do not tend to think that only one of the two clauses of the *Distinct Existence Thesis* is mistaken, perhaps for the reason that Shoemaker (1990) gives: if we prohibit self-contradictions and first-person agnosticism, luminosity and infallibility seem roughly equivalent. For some other reasons to think these two notions go together, see Bilgrami (2006).

33. Even though there are many disagreements in the literature on self-deception, theorists tend to agree that the *explanandum* involves a false higher-order belief. For discussion, see chapter 6 of Fernandez (2013).

have the normative property of commitments" (2006, pg. 315). Similarly, Horgan and Kriegel (2007) defend an infallibility thesis, but they explicitly restrict it to beliefs about the phenomenal properties of one's conscious experiences. So, it is possible for a constitutive theorist to maintain either of the claims denied by the *Distinct Existence Thesis*, just as long as they restrict the scope of their central thesis. They can then insist that cases of either self-ignorance or self-error occur only with respect to psychological kinds that fall outside of some privileged class.[34]

There is a second option available to a constitutive theorist seeking to accommodate self-ignorance or fallibility, which is to argue that there are some additional background conditions on the relevant constitutive relations holding. According to this line of thought, it is not that the attitudes of which we are ignorant or mistaken make up a distinct psychological kind; it is rather that unless certain crucial conditions are met, a subject's psychological attitudes are not necessarily connected to her higher-order beliefs about them. This is the sort of constitutive view associated with Shoemaker, who frequently stresses that one needs rationality and conceptual capacities in addition to first-order attitudes in order to "automatically" have a true second-order belief (1994, pg. 288; cf. 2009). A similar thought is expressed by Jane Heal's claim that a "second-level belief contributes a necessary element to a *set of conditions* which are jointly sufficient for the first-level state" (2001, pg. 5, emphasis added). Someone who holds this sort of constitutive view is thinking of the relation between *a*'s belief that $M(a)$ and $M(a)$ as a sort of conditional necessity. Call the relevant background conditions C. Heal's suggestion is basically that, necessarily, *a*'s believing that $M(a)$ *plus* C is sufficient for $M(a)$, and Shoemaker suggests that $M(a)$ *plus* C entails *a* believing that $M(a)$. On either view, a subject's psychological attitude stands in a necessary constitutive relation to her higher-order belief, but the holding of this relation is conditional on something else — rationality, conceptual sophistication, or whatever. The constitutive theorist can then argue that self-ignorance or error occur only when conditions C fail to hold.[35]

There is no space remaining in this essay to fully discuss the various ways in which a constitutive theorist can try to account for cases of self-ignorance or error. So, it is important to understand that the challenge these cases pose for constitutive theories of self-knowledge is an explanatory one. It is not that constitutive views make either self-ignorance or error impossible; it is rather that they significantly complicate their intelligibility. As we have just seen, there are a number of things a constitutive theorist could say to accommodate *prima facie* counterexamples to her theory. However, the constitutive theorist is forced to turn to comparatively more complicated explanations for instances of self-error or ignorance.[36] By contrast, if we adopt a view that is compatible with the *Distinct Existence Thesis*, much simpler explanations become available. For example, one might pursue the promising idea that we sometimes make mistakes about our beliefs because distraction or fatigue masks our standing disposition (*First-Personal Dispositions*) to form higher-order beliefs about them. This is

34. Some constitutive theorists also reject the notion that self-deception involves an incorrect belief about one's attitudes. For instance, both Bilgrami (2006) and Shoemaker (2009) argue that self-deceived subjects have true beliefs about their attitudes, but also have a belief that is inconsistent with the one they self-ascribe. I assume that there are cases of self-error that are not self-deception, and so a constitutive theorist will need to adopt an additional strategy as well.

35. Although I lack space to discuss the merits of this sort of proposal in detail, I think that one might reasonably question whether conditional necessities are really necessities. If there are some conditions in which an X can fail to stand in a relation to Y, then it seems like the two are not really necessarily connected. They may well stand in that relation in every world in which some further condition C holds, but we might question whether that is sufficient for necessity. For further discussion, see Wilson (2010). Of course, it might still be true that in ordinary psychological conditions, if one *a* is in *M* then *a* will believe that $M(a)$ (or vice versa), but since this isn't necessary it is compatible with the *Distinct Existence Thesis*. The fact that most people know about their own psychological attitudes, or are at least in a position to, is an *explanandum* for a theory of self-knowledge. Non-constitutive theories will presumably need to account for this on the basis of the reliability of certain contingent relations.

36. For an example of the kind of explanatory epicycles I have in mind, see Horgan and Kriegel (2007).

a straightforward causal explanation, but it would lose credibility if we were to accept a constitutive theorist's picture about a how subject's psychological attitudes are necessarily connected to her higher-order beliefs about them. That, it seems to me, is a significant disadvantage for constitutive theories.[37]

One reason the constitutive approach can look attractive is that we are usually not wrong about our psychological attitudes. Most of the time, when everything is working properly, we are not only in a position to know about our own attitudes, but we often do know about them. That is a basic fact about ordinary self-knowledge that any theory must account for. But an equally important truth is that each of us can sometimes be either wrong about or ignorant of the existence or character of our own attitudes. In this essay, I have tried to illustrate how we might accommodate both of these truths. If I am right, one can accept the *Distinct Existence Thesis* while nevertheless acknowledging that first-person self-knowledge has certain distinctive epistemic properties and that mistakes in self-knowledge are very rare.[38]

37. A different objection to constitutive theories that has received a fair amount of attention is they make self-knowledge neither a cognitive nor an epistemic "achievement' (see Boghossian 1989; Fernandez 2013; Fricker 1998; and Peacocke 1999). Obviously if this objection is sound, it is another reason to resist the constitutive approach.

38. Earlier versions of different parts of this essay were presented at a BPhil seminar at Oxford, the Pacific APA, and the Self-Knowledge and Agency Conference at St. Hilda's College, Oxford. On all of these occasions, I was fortunate to receive probing questions and helpful suggestions that contributed to the development of the essay. This essay has also benefited tremendously from many philosophical discussions with the following people: Andreas Anagnostopoulos, Tony Bezsylko, John Campbell, Stanley Chen, David Ebrey, Anil Gomes, Nick Jones, Markus Kohl, Mike Martin, Berislav Marušić, Christopher Peacocke, Sherri Roush, Ian Schnee, Josh Sheptow, James Stazicker, Barry Stroud, Daniel Warren, and Jo Wolff. I am grateful to them. I owe a special note of thanks to Anil Gomes and the late Tony Brueckner, both of whom were thoughtful and generous in formally responding to parts of this essay, and greatly helped me clarify certain ideas. Finally, I would like to thank two anonymous readers for this journal for providing superb comments.

References

Audi, P. (2012): "Grounding: Toward a Theory of the In-Virtue-of Relation". *The Journal of Philosophy* 109 (12): 685–711.

Bilgrami, A. (2010): "*Précis* of Self-Knowledge and Resentment". *Philosophy and Phenomenological Research* 81 (3): 749–765.

Bilgrami, A. (2006): *Self-Knowledge and Resentment*. Cambridge: Harvard University Press.

Boghossian, P. (1989): "Content and Self-Knowledge". In Peter Ludlow and Norah Martin (eds.), *Externalism and Self-Knowledge*. Stanford: CSLI Publications.

Bortolotti, L. (2010): *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

Boyle, M. (2016): "Additive Theories of Rationality: A Critique". *European Journal of Philosophy* 24 (3): 527–555.

Boyle, M. (2011): "Transparent Self-Knowledge". *Aristotelian Society Supplementary Volume* 85 (1): 223–241.

Boyle, M. (2009): "Two Kinds of Self-Knowledge". *Philosophy and Phenomenological Research* 78 (1): 133–164.

Burge, T. (1998): "Reason and the First Person". In Crispin Wright, Barry C. Smith and Cynthia Macdonald (eds.), *Knowing Our Own Minds*. Oxford: Oxford University Press.

Burge, T. (1996): "Our Entitlement to Self-Knowledge". *Proceedings of the Aristotelian Society* 96: 91–116.

Byrne, A. (2011): "Transparency, Belief, Intention". *Aristotelian Society Supplementary Volume* 85 (1): 201–221.

_____. (2005): "Introspection". *Philosophical Topics* 33 (1): 79–104.

Cassam, Q. (2014): *Self-Knowledge for Humans*. Oxford: Oxford University Press.

Coliva, A. (2012): "One Variety of Self-Knowledge: Constitutivism and Constructivism". In Annalisa Coliva (ed.), *The Self and Self-Knowledge*. Oxford: Oxford University Press. (2009): "Self-Knowledge and Commitments". *Synthese* 171 (3): 365–375.

Fernández, J. (2013): *Transparent Minds: A Study of Self-Knowledge*. Oxford: Oxford University Press.

Fine, K. (1995): "Ontological Dependence". *Proceedings of the Aristotelian Society* 95 (1995): 269–290.

Finkelstein, D. (1999): "On Self-Blindness and Inner Sense". *Philosophical Topics* 26 (1/2): 105–119.

Fricker, E. (1998): "Self-Knowledge: Special Access versus Artefact of Grammar—A Dichotomy Rejected". In Crispin Wright, Barry C. Smith and Cynthia MacDonald Macdonald (eds.), *Knowing Our Own Minds*. Oxford: Oxford University Press.

Friedman, J. (2013): "Rational Agnosticism and Degrees of Belief". In Tamar Szabó Gendler and John Hawthorne (eds.), *Oxford Studies in Epistemology*, *Volume 4*. Oxford: Oxford University Press.

Gendler, T. S. (2008): "Alief and Belief". *The Journal of Philosophy* 105 (10): 634–663.

Handfield, T. and Bird, A. (2008): "Dispositions, Rules, and Finks". *Philosophical Studies* 140 (2): 285–298.

Heal, J. (2001): "On First-Person Authority". *Proceedings of the Aristotelian Society* 102 (1): 1–19.

Hieronymi, P. (2005). "The Wrong Kind of Reason". *The Journal of Philosophy* 102 (9): 437–457.

Horgan, T. and Kriegel, U. (2007): "Phenomenal Epistemology: What Is Consciousness that We May Know It So Well?". *Philosophical Issues* 17 (1): 123–144.

Hume. D. 1741. *A Treastise of Human Nature*. D. Norton (ed.). Oxford: Oxford University Press.

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., and Hardin, C. D. (2009): "The Existence of Implicit Bias Is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies that No Manager Should Ignore". *Research in Organizational Behavior* 29: 39–69.

Kind, Amy (2003): "Shoemaker, Self-Blindness and Moore's Paradox". *Philosophical Quarterly* 53 (210): 39–48.

Kornblith, H. (2012): *On Reflection*. Oxford: Oxford University Press.

Martin, M. G. F. (1998): "An Eye Directed Outward". In Crispin Wright, Barry C. Smith and Cynthia Macdonald (eds.), *Knowing Our Own Minds*. Oxford: Oxford University Press.

Moran, R. (2012): "Self-Knowledge, 'Transparency', and the Forms of Activity". In Declan Smithies and Daniel Stoljar (eds.), *Introspection and Consciousness*. Oxford: Oxford University Press.

_____. (2003): "Responses to O'Brien and Shoemaker.' *European Journal of Philosophy* 11 (3): 402–419.

_____. (2001): *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

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

_____. (1998): "Conscious Attitudes, Attention, and Self-Knowledge". In Crispin Wright, Barry C. Smith and Cynthia Macdonald (eds.), *Knowing Our Own Minds*. Oxford: Oxford University Press.

Rayo, A. (2013): *The Construction of Logical Space*. London: Oxford University Press.

_____. (2009): "Towards a Trivialist Account of Mathematics". In Otávio Bueno and Øystein Linnebo (eds.), *New Waves in Philosophy of Mathematics*. City?: Palgrave MacMillan.

Rödl, S. (2007): *Self-Consciousness*. Cambridge: Harvard University Press.

Rosen, G. (2010): "Metaphysical Dependence: Grounding and Reduction". In Bob Hale and Aviv Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press.

Rumfitt, I. (1994): "Frege's Theory of Predication: An Elaboration and Defense, with Some New Applications". *The Philosophical Review* 103 (4): 599–637.

Scanlon, T. M. (1998): *What We Owe to Each Other*. Cambridge: Harvard University Press.

Schwitzgebel, E. (2008): "The Unreliability of Naïve Introspection". *Philosophical Review* 117 (2): 245–273.

_____. (2002): "A Phenomenal, Dispositional Account of Belief". *Noûs* 36 (2): 249–275.

Shah, N. and Velleman, J. D. (2005): "Doxastic Deliberation". *The Philosophical Review* 114 (4): 497–534.

Sider, T. (2007): "Parthood". *The Philosophical Review* 116 (1): 51–91.

Silins, N. (2012): "Judgment as a Guide to Belief". In Declan Smithies and Daniel Stoljar (eds.), *Introspection and Consciousness*. Oxford: Oxford University Press.

Shoemaker, Sydney (2009): "Self-Intimation and Second-Order Belief" *Erkenntnis* 71 (1): 35–51.

_____. (1996): "Moore's Paradox and Self-Knowledge". In his *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

_____. (1994): "Self-Knowledge and 'Inner Sense'". In his *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

_____. (1990): "First-Person Access". In his *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

_____. (1988): "On Knowing One's Own Mind". In his *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.

Snowdon, P. (2012): "How to Think About Phenomenal Self-Knowledge". In Annalisa Coliva (ed.), *The Self and Self-Knowledge*. Oxford: Oxford University Press.

Srinivasan, A. (2015): "Are We Luminous?". *Philosophy and Phenomenological Research* 90 (2): 294–319.

Sturgeon, S. (2015): "The Tale of Bella and Creda". *Philosophers' Imprint* 15 (31): 1–9.

Williamson, T. (2008): "Why Epistemology Can't be Operationalized". In Quentin Smith (ed.), *Epistemology: New Essays*. Oxford: Oxford University Press.

_____. (2000): *Knowledge and Its Limits*. Oxford: Oxford University Press.

Wilson, J. (2010): "What Is Hume's Dictum, and Why Believe It?". *Philosophy and Phenomenological Research* 80 (3): 595–637.

Wright, C. (2012): "Replies". In Crispin Wright and Annalisa Coliva (eds.), *Mind, Meaning, and Knowledge: Themes from the Philosophy of Crispin Wright*. Oxford: Oxford University Press.

_____. (1998): "Self-Knowledge: The Wittgensteinian Legacy". In Crispin Wright, Barry C. Smith and Cynthia Macdonald (eds.), *Knowing Our Own Minds*. Oxford: Oxford University Press.

_____. (1989): "Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention". *The Journal of Philosophy* 86 (11): 622–624.

Zimmerman, A. (2006): "Basic Self-Knowledge: Answering Peacocke's Critiques of Constitutivism" *Philosophical Studies* 128 (2): 337–379.