

Self-Knowledge and Rational Agency: a defense of empiricism

Brie Gertler, University of Virginia

Forthcoming in Philosophy and Phenomenological Research.

This is the penultimate version. Please use published version for citations.

How does one know one's own beliefs, intentions, and other attitudes? Many responses to this question are broadly *empiricist*: they take self-knowledge to be epistemically based in empirical justification (perhaps consisting in introspective evidence) or warrant (perhaps deriving from reliable inference or inner sense). Empiricist approaches to self-knowledge face an influential objection: that empiricism portrays us as mere observers of a passing cognitive show, and thereby neglects the fact that believing and intending are things we *do*, for *reasons*. Critics who press this objection contend that our capacity for self-knowledge derives from our rational agency—our ability to conform our attitudes to our reasons, and to commit ourselves to those attitudes through avowals (Burge 1996; Moran 2001; Bilgrami 2006; Boyle 2009). I'll refer to this position as *agentialism* about self-knowledge.¹

This paper has two goals. The first is exegetical: to identify agentialism's defining thesis and precisely formulate the agentialist challenge to empiricism. The second goal is to defend empiricism from the agentialist challenge. I propose a way to understand the role of agency in reasoning and avowals, one that does justice to what is distinctive about these phenomena yet is compatible with empiricism about self-knowledge.

Section 1 spells out the core dispute between empiricism and agentialism, arguing that it concerns the epistemic basis for self-knowledge rather than the method by which self-knowledge is achieved. Section 2 explicates the agentialist challenge to empiricism and distinguishes it from a related challenge (Shoemaker 1994). Section 3 begins my empiricist response to the agentialist challenge. I propose that the agential features that characterize avowals and critical reasoning about one's attitudes are present in other types of agency as well. These features derive from a subject's occupying what I call the *agential position*: the position from which she can act directly on the basis of her reasons. Section 4 argues that my proposal makes sense of the agency involved in critical reflection on our attitudes, and Section 5 argues that it makes sense of the agency associated with avowals. Section 6 shows that one can occupy the agential position on an attitude even if one's knowledge of that attitude is purely empirical. So the empiricist conception of self-knowledge can do justice to the rational agency phenomena. Empiricism thus withstands the agentialist challenge.

¹ In previous work, I referred to this position as "rationalism" (Gertler 2011). But I now favor the label "agentialism", as it allows for a distinction between this position and those that invoke our rationality in explaining self-knowledge without tying rationality to agency (such as Shoemaker 1994 and Gallois 1996).

1. The Dispute between Empiricism and Agentialism

Historically, most philosophers have held that knowledge of one's own mental states is achieved through introspective observation. This view is clearly a form of empiricism about self-knowledge. But some forms of empiricism are non-introspectivist. Dretske's "displaced perception" account, which is plainly empiricist, epistemically grounds self-knowledge in *perceptual* rather than introspective observation.² An account needn't even construe self-knowledge as observational, to qualify as empiricist. Some empiricists take self-knowledge to consist in self-attributions of attitudes that are empirically warranted by virtue of a reliable link with the attitudes themselves, where the process by which one arrives at self-knowledge need not involve *observation* in any ordinary sense (Byrne 2011, Fernández 2013).

Agentialists agree that knowledge of one's own attitudes sometimes rests on purely empirical justification or warrant. But on their view this kind of self-knowledge is deviant: a thinker who *detects* a belief or intention, even through a uniquely first-personal method like introspection, is alienated from that attitude. She regards her attitude as a fixed feature of the internal landscape; and her relation to the attitude is merely that of an observer, albeit one who is especially well-situated. Agentialists charge that empiricist accounts of self-knowledge neglect the truly distinctive relation between a thinker and her attitudes, a relation involving rational agency.

The agentialist challenge to empiricism centers on beliefs and intentions, since these attitudes are answerable to reasons and hence taken to express rational agency. "Our rational beliefs and intentions are not mere mental attitudes, but active states of normative commitment" (Korsgaard 2009, 39). Agentialists acknowledge that some beliefs and intentions are insensitive to reasons, and they grant that these may be knowable only through observation. Agentialism's central objection to empiricism is that empiricism assimilates all self-knowledge of the attitudes to self-knowledge of reasons-insensitive attitudes like these.

At first blush, agentialism seems not to be a competitor to empiricist accounts of self-knowledge. Empiricist accounts are principally concerned with how attitudes are *known*: they seek to explain what justifies or warrants our self-attributions. Knowledge ordinarily involves the knower (specifically, her beliefs) being appropriately controlled by what is known. E.g., my beliefs about my dog's whereabouts qualify as knowledge only insofar as they are appropriately controlled by his actual location. But our agency, relative to an event or situation, concerns a direction of control that is the reverse of the direction ordinarily operative in knowledge. Agency is a matter of the agent's *controlling* an event or situation. The agentialist emphasis on agency thus seems to change the topic, from how we *know* our own attitudes to how we *shape* those attitudes.³

² According to Dretske, self-knowledge is "a form of perceptual knowledge that is obtained—indeed, can only be obtained—by awareness of non-mental objects" (Dretske 1994, 264).

³ The idea that knowledge can involve the control of what is known is familiar from Anscombe (1963), who borrows Aquinas' phrase "the cause of what it understands" to describe how my knowledge of what I'm doing (or intending) *makes it the case* that I'm doing that thing, rather than something extensionally equivalent. My point here is only that in focusing on how we *control* rather

Doubts about whether agentialism is an epistemic theory are bolstered by some agentialists' apparent dismissal of epistemic concerns. In his influential 2001 book, Moran argues that the most remarkable cases of self-knowledge are those in which one *avows* a belief or intention, on the basis of reasons bearing on it. But although that book bears the subtitle "an essay on self-knowledge", Moran doesn't explain there how this procedure provides an epistemic basis for the resulting self-attribution. Moran's treatment of the epistemic, in that book, is mainly focused on discounting its contribution to the distinctiveness of self-knowledge.

We may allow any manner of inner events of consciousness, any exclusivity and privacy, any degree of privilege and special reliability, and their combination would not add up to the ordinary capacity for self-knowledge. For the connection with the *avowal* of one's attitudes would not be established by the addition of any degree of such epistemic ingredients. (*ibid.*, 93)

Remarks like this, together with the emphasis on a direction of control that is the reverse of that usually operative in knowledge, promote the impression that agentialism is not an epistemic position. On an alternative interpretation, the agentialist objection to empiricism consists in the charge that empiricism overrates the significance of the thinker's epistemic access to her attitudes, and correspondingly underrates the significance of the thinker's rational agency relative to her attitudes. After all, the agentialist might say, what makes my beliefs and intentions truly *mine* is the fact that they express my commitments, for which I am responsible. That I have privileged epistemic access to them is much less important.

This interpretation of the debate, while appealing, is inadequate. For one thing, it construes empiricism as committed to a claim on which that approach is neutral. Empiricism is a view about the epistemic basis for self-knowledge: that the epistemic basis for knowledge of the attitudes is empirical justification or warrant. Empiricism is neutral as to whether epistemic relations are generally more important than agential relations, and whether they explain what makes one's attitudes one's own.

More to the point, agentialism *is* an epistemic view. Burge is especially clear about this. He argues that, as rational agents, we are epistemically entitled to "a certain range of judgments about our own thoughts and attitudes." (Burge 1996, 91) Moran also plainly regards his own agentialism as an epistemic position. When pressed by critics about the epistemic basis for self-knowledge, Moran responded by proposing that our right to assume that we are rational agents gives us the right to self-attributions arrived at through engagement with reasons (Moran 2003). Agentialism takes self-knowledge to rest on epistemic entitlements or rights deriving from rational agency. It is thus a direct competitor to empiricism, regarding the epistemology of self-knowledge.

Another tempting, but ultimately inadequate, interpretation of this debate construes empiricism and agentialism as rival views about the *method* by which self-

than *are controlled by* the object of knowledge, the agentialist may appear to be addressing a question different from the question addressed by empiricist accounts.

knowledge is achieved. Some proposed methods of achieving self-knowledge will be more amenable to one of these camps than to the other. But the crux of the dispute between empiricists and agentialists does not concern the particular process by which we arrive at self-knowledge. To see this, consider the “transparency method”, whereby directly considering an issue provides knowledge of what one believes about that issue. In Evans’ famous example, I respond to the question “Do you believe that there will be a third world war?” by considering whether there will be a third world war (Evans 1982). Moran endorses this method, giving it an agentialist construal: as a rational agent, he says, one has the “right to assume” that one can answer a question about one’s beliefs by use of this method (Moran *ibid*). But some empiricist philosophers endorse the transparency method on the grounds that its use satisfies empiricist requirements for knowledge, concerning reliability or safety (Byrne 2011, Fernández 2013). So the principal dispute between empiricists and agentialists does not concern the *method* by which self-knowledge is achieved.

The dispute between empiricism and agentialism about self-knowledge is, at its core, a dispute about the *epistemic basis* for knowledge of one’s own attitudes. For empiricists, the epistemic basis for self-knowledge is empirical justification or warrant. Depending on the particular empiricist account at issue, one’s grasp of one’s own attitude constitutes self-knowledge insofar as it is based in appropriate evidence or reasons; or the result of a reliable or properly functioning belief-forming process; or generated by the exercise of an epistemic virtue; or couldn’t easily have been false; etc. Agentialists maintain that the epistemic basis for paradigm instances of self-knowledge is not (primarily) empirical justification or warrant. Instead, judgments about our attitudes are epistemically based in an *entitlement* that derives from our rational agency, and that is irreducible to empirical justification or warrant.

Commitment to the following thesis thus qualifies a view as agentialist.

Agentialism: In centrally significant cases, knowledge of one’s own attitudes is epistemically based in an *entitlement* that derives from one’s rational agency relative to one’s attitudes. This entitlement is irreducible to the kinds of epistemic factors (evidence, reliability, safety, etc.) countenanced by empiricism.

The sense of “empiricism” here is, I hope, tolerably clear. The key point is that the position I’m calling “agentialist” (versions of which are defended by Bilgrami, Boyle, Burge, and Moran) takes the fact of our rational agency to epistemically contribute to self-knowledge in a way that empiricist philosophers (such as those mentioned above) deny that it does. The core dispute between empiricism and agentialism is not about what makes an attitude one’s own, or about the method used to achieve self-knowledge. It is about the epistemic basis for self-knowledge.

2. The Agentialist Challenge to Empiricism

2.1 *Transcendentally derived entitlement*

I will now explicate the agentialist challenge to empiricism. To keep the scope of my exegesis manageable, I will focus on the versions of agentialism advanced by Burge

and Moran. Both of these authors use transcendental reasoning to show that our rational agency generates an entitlement (or “right”) that epistemically grounds self-knowledge.⁴ The use of transcendental reasoning is perhaps unsurprising, given that both of these authors explicitly register debts to Kant. Because epistemic issues figure more prominently in Burge’s discussion, my discussion will largely focus on his argument.

But I begin by briefly sketching Moran’s transcendental argument. This argument explains the basis for the transparency method, by reference to Moran calls “the Transparency Condition”: that one’s beliefs suitably reflect (are “transparent” to) one’s evidence, that is, one’s reasons for belief. On his view, it is essential to our conception of ourselves as rational agents that we take the Transparency Condition to apply to our beliefs.

[I]t is only because I assume that what I actually believe about X can be determined, made true by, my reflection on X itself, that I have the right to answer a question about my belief in a way that respects the Transparency Condition. ... So I take the simple Transparency condition to have the surprising consequence that it brings us up to the region of something like a Transcendental assumption of Rational Thought as it has figured elsewhere in Kantian and post-Kantian philosophy, as venerable and familiar as it is obscure. (Moran 2003, 406)

Moran’s reasoning is as follows. As a rational agent, I have the right to regard my beliefs as transparent to my reasons for belief. I wouldn’t have this right unless I had the right to adopt the “Transcendental assumption of Rational Thought”: that is, to assume that my deliberations, in which my reasons are operative, *shape* my beliefs. It is this “Transcendental assumption” that entitles me to answer a question as to my beliefs about X by considering X itself—that is, by engaging with *reasons* for belief. So the self-attributions generated by the transparency method qualify as knowledge (at least partly) because of an entitlement derived from the Transcendental assumption of Rational thought.

Burge’s transcendental argument, while similar in spirit, diverges from Moran’s in some details. Burge derives our epistemic entitlement to self-attributions specifically from *obligations* we bear as rational agents. As rational thinkers, he says, we are obligated to (try to) satisfy certain rational norms: e.g., that beliefs should conform to evidence. Working to satisfy these norms requires relying on judgments about our attitudes. E.g., to assess my beliefs for conformance with my evidence, I must know what it is that I

⁴ While these views are broadly similar in outlook, they differ in some details. For Burge, the responsibility associated with rational agency plays a pivotal role in generating epistemic entitlement. Moran appears to hold that rational agency generates epistemic rights directly, independent of responsibility, although he allows that a sense of responsibility for one’s attitudes makes one *aware* that one is a rational agent. Another difference is that Moran is committed to the claim that we achieve self-knowledge through the transparency method. Burge does not commit himself to any particular view about how self-knowledge is achieved; and there are reasons to doubt that Burge’s view is compatible with Moran’s claims on behalf of transparency. (For details, see Gertler 2011, ch. 6.) These differences do not bear directly on the current discussion.

believe. On Burge's account, our responsibility to satisfy rational norms entitles us to those judgments about our attitudes that are crucial for satisfying those norms.

[I]f one lacked entitlement to judgments about one's attitudes, one could not be subject to rational norms governing how one ought to alter those attitudes given that one had reflected on them. (Burge 1996, 101)

For Burge, entitlements are "epistemic warrants that need not be conceptually accessible to the warranted individual" (Burge 2003, 521). Entitlement appears to involve two factors: reliability and permissibility. Judgments to which one is entitled are reliable, that is, "well-positioned to achieve or indicate truth in normal circumstances" (*ibid.*, 507). It is the second factor involved in entitlement, permissibility, that distinguishes entitlement from standard externalist warrant. A subject is not entitled to a particular judgment unless that judgment is epistemically permissible for her. What makes the relevant judgments about our own attitudes permissible, on Burge's view, is that they are required for the kind of critical self-reflection that we are obligated to undertake, as rational agents. The nature and source of this permissibility can be grasped only through transcendental reflection on our rational nature.

2.2 *The agentalist challenge to empiricism: initial formulation*

Both proponents and critics of agentalism tend to sketch the argument for agentalism in ways that are suggestive but imprecise. Here is a formulation that initially seems to capture Burge's reasoning (in the passage from Burge 1996 just quoted). I will amend this formulation below.

1. As rational agents, we are obligated to satisfy certain rational norms.⁵
2. Satisfying some of those rational norms requires reasoning critically about one's attitudes.
3. One cannot reason critically about one's attitudes unless one can know one's attitudes.
4. We can satisfy our obligations, as rational agents, only if we can know our attitudes. (from 1-3)
5. "*S* is obligated to *A*" implies "*S* can *A*".
6. As rational agents, we can know our attitudes. (from 4 & 5)

Thus formulated, the argument resembles Shoemaker's famous "self-blindness" argument against a broad class of empiricist theories of self-knowledge (Shoemaker 1994).⁶ The self-blindness argument is driven by the claim that one is not a fully rational

⁵ The extent of our obligations here is not clear. Presumably, the obligation to satisfy rational norms doesn't require us to be *ideal* reasoners. But the precise extent of our obligations doesn't matter for this argument. What matters is that rational norms impose certain obligations, obligations that can be met only through achieving self-knowledge.

⁶ Shoemaker is not committed to the agentalist thesis. However, his 1994 arguments challenging the "inner sense" model of self-knowledge are a principal inspiration for alternatives to empiricism that

being unless one can revise one's attitudes when appropriate. This in turn requires that one is capable of self-knowledge—or, in Shoemaker's words, "sensitive to the contents of one's belief-desire system".

[I]t is essential to being a rational being that one be 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. (Shoemaker *ibid.*, 285)

Shoemaker concludes that the capacity for self-knowledge is not a merely contingent feature of rational thinkers, on a par with vision or other perceptual capacities. For while some rational creatures are visually blind, no rational creature (possessing the suitable repertoire of concepts) could be *self-blind*, that is, incapable of knowing its own mental states. Shoemaker's direct target here is the claim that self-knowledge occurs through a quasi-perceptual "inner sense". But he takes his conclusion, that rational creatures necessarily possess the capacity for self-knowledge, to threaten a wide range of empiricist accounts of self-knowledge.

This conclusion does not, however, express the challenge to empiricism that concerns us here. For it does not say anything about the epistemic basis for self-knowledge. As Kind (2003) demonstrates, the conclusion that rational creatures must be capable of self-knowledge does not limit the possible *routes* to self-knowledge. So it is entirely compatible with empiricism.

The empiricist can agree that, necessarily, a rational thinker will be capable of self-knowledge. This is the Armstrong's position. Armstrong, a staunch empiricist, posits a connection between reasoning ("purposive trains of mental activity") and self-knowledge similar to that described by Shoemaker.

If there are to be purposive trains of mental activity, then there must equally be some means by which we become apprised of our current mental state. Only so can we adjust mental behaviour to mental circumstances. ... So there must be a way of becoming aware of our current mental state, which means that there must be introspection. (Armstrong 1968/1993, 327)

Armstrong is an inner sense theorist: he construes self-knowledge as the result of "a self-scanning process in the brain" (*ibid.*, 324). He can explain the impossibility of self-blindness as follows: no thinker qualifies as *rational* unless these self-scanning processes occur when appropriate, and are generally reliable. That one is rational thus entails that one has the capacity for self-knowledge, since having an appropriately functioning "self-scanner" is a *component* of (or condition for) rationality. So the empiricist can accept the

emphasize the importance of rationality, including agentialism. I critically assess those arguments in Gertler 2011, ch. 5.

statement “Necessarily, if *S* is rational, then *S* is capable of self-knowledge”, by construing the necessity as *de dicto*.⁷

Shoemaker’s self-blindness argument implies that rational thinkers must be capable of self-knowledge. But it does not directly threaten empiricism.

To construct a threat to empiricism, the agentialist must further argue that self-knowledge epistemically grounded in purely empirical justification or warrant—I’ll call this *empirical self-knowledge*—cannot make the needed contribution to rational thought. This is the burden of the amended formulation of the transcendental argument, which better captures Burge’s reasoning. The modified premises are marked with an asterisk. The primary change is in the third premise.

2.3 *The agentialist challenge to empiricism: amended formulation*

1. As rational agents, we are obligated to satisfy certain rational norms.
2. Satisfying some of those rational norms requires reasoning critically about one’s attitudes.
- 3*. One cannot reason critically about one’s attitudes unless one can know one’s attitudes *non-empirically*.
- 4*. We can satisfy our obligations, as rational agents, only if we can know our attitudes *non-empirically*. (from 1-3*)
5. “*S* is obligated to *A*” implies “*S* can *A*”.
- 6*. As rational agents, we can know our attitudes *non-empirically*. (from 4* & 5)

To illustrate the inadequacy of empirical self-knowledge for critical self-reflection—that is, to support (3*)—Burge presents the following example. A juror knows that he believes that the defendant is guilty. Reflection leads him to realize that his evidence for the defendant’s guilt is very weak. Although the juror recognizes that he therefore has reason to suspend the belief, the belief perseveres. The juror then “must face a question of how, by what means, to make [his] reasons [for suspending the belief] effective”. So this process “is not critical reasoning” (Burge 1996, 113). In genuine critical reasoning about a belief, the realization that the belief is unjustified eliminates it *immediately*. More generally, beliefs are *directly* sensitive to reasons. (Moran agrees: “...it is because the deliberator declares the authority of reason over his thought and action that

⁷ An agentialist might object that my response to Shoemaker misconstrues the modal relation between rationality and self-knowledge. My response shows how the empiricist can accommodate the *de dicto* necessity of the statement “If *S* is rational, then *S* is capable of self-knowledge”. But Shoemaker sees this as a matter of *de re* necessity: he says that the capacity for self-knowledge is “essential to being a rational being”. Some empiricists will regard this kind of *de re* necessity with suspicion. However, our concern here is with empiricism as a view about self-knowledge; and empiricism in that sense is neutral about modality and identity conditions. So there is nothing to bar an empiricist from granting that the capacity for self-knowledge is a *de re* necessary feature of rational creatures. The empiricist could argue as follows: *being rational* is a *de re* necessary feature of any creature that has it; the capacity for self-knowledge is a component of rationality; so the capacity for self-knowledge is a *de re* necessary feature of rational creatures.

at the conclusion of his thinking there is no further thing he does to make that conclusion his actual belief or his intention...”(Moran 2001, 131.)

On Burge’s diagnosis, the perseverance of the juror’s belief indicates that his reflections are detached from their object: he knows the belief through observation, and hence has only *empirical* self-knowledge. Empirical self-knowledge cannot appropriately contribute to critical reasoning about an attitude because such knowledge involves a gap between the perspective from which the attitude is known and evaluated (“the reviewing point of view”) and the perspective of the attitude itself (“the reviewed perspective”).

If the relation between the reviewing point of view and the reasons or assumptions being reviewed always fit the simple observational model [8], there would never be an immediate rationally necessary connection between justified rational evaluation within the review, on one hand, and its being *prima facie* reasonable within the reviewed perspective to shape attitudes in accord with that evaluation, on the other. ... (The attitudes reviewed would be to the reviews as physical objects are to our observational judgments. They would be purely “objects” of one’s inquiry, not part of the perspective of the inquiry.) (*ibid.*, 109-110)

For Burge, two attitudes belong to a single “perspective” if, and only if, there is an “immediate rationally necessary connection” between them. By contrast, in empirical self-knowledge there are only “brute, contingent, non-rational relations” between the reviewing and reviewed perspectives (*ibid.*, 105). This is why empirical self-knowledge is similar to observational knowledge of external objects.

Burge’s argument for (3*) is now clear. “In critical reasoning, ... one’s relation to the known attitudes is rationally immediate: they are part of the perspective of the review itself.” (*ibid.*, 113) Insofar as the juror’s grasp of his belief that the defendant is guilty is only *empirical* self-knowledge, it is related to the belief only by “brute, contingent, non-rational relations”. So the realization that the belief is poorly justified cannot dislodge it with the rational immediacy required for genuine critical reasoning. Self-knowledge of attitudes that contributes to critical reasoning about one’s own attitudes—self-knowledge that contributes to what I’ll call *critical self-reflection*—must be non-empirical.

3. The Agential Position

The agentialist challenge to empiricism proceeds in two steps. First, the agentialist argues that there is an important difference between treating an attitude as an exercise of

⁸ While Burge uses “the simple observational model” as a label for the view he criticizes, his explication of that view makes it clear that his target is empiricism generally. “The fundamental claim [of the simple observational model] is that one’s epistemic warrant for self-knowledge always rests partly on the existence of a pattern of veridical, but brute, contingent, non-rational relations—which are plausibly always causal relations—between the subject matter (the attitudes under review) and the judgments about the attitudes.” (Burge *ibid.*, 105) It’s worth noting that empiricist views need not construe the relation between an attitude and a judgment as *causal*. The claim that one knows one’s own attitude through direct acquaintance is plainly a version of empiricism. Moreover, on at least some views knowledge by acquaintance is *observational* knowledge.

one's own agency and simply observing the attitude. Second, the agentialist contends that empiricism cannot accommodate this difference, since empiricism assimilates all self-knowledge to the latter, merely observational grasp of attitudes.

The empiricist can object to either step of the agentialist challenge. However, I think the empiricist should accept the first step. For the difference it posits is a genuine one. Treating an attitude as a commitment for which one is responsible does differ from simply registering its presence. The empiricist should do justice to this difference.

I think that the empiricist *can* do justice to this difference, and so I reject the second step of the agentialist challenge. In what follows, I propose an understanding of the distinctively agential dimension of critical self-reflection (critical reasoning about one's own attitudes) and avowals, one that is compatible with an empiricist epistemology of self-knowledge.

Let us examine more closely the difference between critical self-reflection and the kind of process just described, in which beliefs initially resist the force of reasons. For illustrative purposes, I will use an extension of Burge's juror example. My extension involves two jurors.

- (1) *The deliberative juror.* The deliberative juror grasps that she believes that the defendant is guilty. She then reviews the prosecution's case, and comes to realize that her evidence for the defendant's guilt is weak. This realization directly results in the belief's suspension.
- (2) *The detached juror.* The detached juror grasps that he believes that the defendant is guilty. He then reviews the prosecution's case, and comes to realize that his evidence for the defendant's guilt is weak. Although he regards this as ample reason to suspend the belief, his belief survives this realization: he notices that he is still strongly inclined to think of the defendant as guilty. (He might even understand this as a case of belief perseverance, and chalk it up to bias.) He tries to rid himself of the belief by imagining the defendant doing what he claims he was doing at the time of the crime—sleeping in bed, say. By repeatedly engaging in this imaginative exercise, the juror eventually shakes his belief in the defendant's guilt.

For both jurors, the weakness of the evidence provides a reason to suspend the belief. The difference between how the deliberative juror acts on this reason, and how the detached juror acts on it, reflects a more general variation in ways that reasons guide actions. Sometimes one responds to a reason to φ directly by φ -ing. At other times, one responds by directing one's attention and efforts towards a discrete task aimed at bringing it about that one φ -s.

For an example of non-cognitive action guided by reasons via the latter, less direct route, consider the process of adapting to a prosthetic limb. Some modern prosthetics operate by detecting nerve signals in the brain, which trigger a robotic mechanism. But the method of using thought to control these prosthetic limbs is far from intuitive.

Replicating their actions with robotic arms can be excruciatingly difficult, requiring amputees to understand the distinct muscle contractions

involved in movements they once did without thinking. To bend the elbow, for instance, requires thinking about contracting a biceps, though the muscle no longer exists. But the thought still sends a nerve signal that can tell a prosthetic arm to flex. Every action, from grabbing a cup to turning the pages of a book, requires some such exercise in the brain. (Dao 2012)

Dao describes the experience of Marine Corporal Sebastian Gallegos, who lost his arm to an I.E.D. in Afghanistan, as he struggles to pick up a sponge with his new prosthetic.

Close hand, raise elbow, he says to himself. The mechanical arm rises, but the claw-like hand opens, dropping the sponge. Try again, the therapist instructs. Same result. Again. Tiny gears whir, and his brow wrinkles with the mental effort. The elbow rises, and this time the hand remains closed. He breathes. Success. (*ibid.*)

Suppose that Gallegos accepts the principle *one should salute one's superior officers*. Seeing his commanding officer enter the room gives Gallegos a reason to raise his arm in salute. At this early stage of learning to use the new arm, this reason will generate Gallegos' raising of his arm only indirectly. He must first consider *how* to raise his arm. Like the juror with the persevering belief, for a moment he "face[s] a question of how, by what means, to make [his] reasons effective" (Burge *ibid.*, 113). He remembers that the way to bend the arm is to think about flexing his biceps. He does this and, after a brief struggle, he salutes.

Now imagine Gallegos shortly before his injury. Seeing his commanding officer enter the room generates a reason for Gallegos to raise his arm in salute. This reason issues in action directly, since he doesn't have to think about *how* to do this. He simply salutes.

The contrast between these two scenarios centers on how Gallegos acts on his reason to raise his arm. In the prosthetic scenario, he does what he has reason to do *indirectly*, by turning his attention to a discrete task (thinking about flexing his biceps) aimed at bringing it about that he does what he has reason to do. In the pre-injury scenario he directly does what he has reason to do. The contrast between these two scenarios is analogous, in this respect, to the contrast between the two juror scenarios. The detached juror does what he has reason to do *indirectly*, by turning his attention to a discrete task (imagining the defendant sleeping at the time of the crime), whereas the deliberative juror directly does what she has reason to do.

My analysis of critical self-reflection and avowals centers on the ability to directly do what one has reason to do. One who has this ability occupies what I call *the agential position*.

S occupies the agential position on φ -ing, relative to reasons R to $\varphi =df.$ S can φ directly on the basis of R .⁹

While still learning to adapt to the prosthetic, Gallegos does not occupy the agential position on saluting (relative to any reasons). If he becomes sufficiently adept at using the prosthetic—if the means of raising his arm becomes “second nature” to him—he will once again occupy that agential position. Of the two jurors, only the deliberative juror occupies the agential position on revising her belief, relative to the weakness of her evidence.

In the next two sections, I use the notion of the agential position to identify the agential dimension of critical self-reflection and avowals, respectively. In section 6 I show how this construal of critical self-reflection and avowals enables the empiricist to answer the agentialist challenge.

4. Critical Self-Reflection

I propose that we understand critical self-reflection as the preservation or revision of attitudes directly on the basis of appropriate reasons—reasons corresponding to rational norms, such as *beliefs should conform to evidence*. In my sense of “agential position”, then, those engaged in critical self-reflection occupy the agential position on preserving or revising attitudes, relative to reasons corresponding to rational norms.

A skeptic about mental agency might contest this proposal, by maintaining that belief revision is something that thinkers simply undergo, rather than an action they perform. On this view, in suspending their beliefs the jurors are not acting on the basis of reasons.

Defending a robust conception of mental agency is no part of my project. My goal is to show that the empiricist can withstand the agentialist’s challenge. That challenge is premised on the claim that critical self-reflection involves a kind of agency. So I grant that claim for current purposes. To deny it would be to dismiss the agentialist challenge outright.

In allowing that not only bodily behavior but also cognitive activities can constitute *acting on the basis of reasons*, I intend this phrase to carry minimal connotations (while remaining compatible with more substantive claims, e.g., that something qualifies as a *reason* to φ only if the agent endorses or would endorse it as such). The specification of the agential position inherits this neutrality. So to say that one occupies the agential position on revising an attitude, relative to certain reasons, is not to say that she *chooses* to act on those reasons, or that she is capable of resisting their force.

⁹ “ S occupies the agential position on...” introduces an opaque context. I may occupy the agential position on *raising my left arm*, and not occupy that position on *raising the most freckled limb in the room*, even if my left arm is the most freckled limb in the room. I may have reason to raise the most freckled limb (if you offer me \$100 for raising the most freckled limb), yet be unable to act directly on the basis of that reason, e.g. if I falsely believe that my right arm is more freckled than my left.

Accounts of action vary as to the conditions they place on *acting on the basis of reasons*. At one end of the spectrum are simple causal accounts, which imply that my dog acts on the basis of a reason when he responds to an itch by scratching. On a slightly more demanding conception, the agent ϕ -s on the basis of reasons R only if she is generally disposed to respond to R by ϕ -ing and she would endorse R as a reason to ϕ if she were to consider the question. More demanding conceptions add further requirements, e.g.: that the agent actually endorses R as a reason to ϕ , that she identifies with the desire to ϕ given R , or that ϕ -ing on the basis of R satisfies some objective norm. I will not attempt to adjudicate among these conceptions of action here. But it's worth noting that the detached juror may well meet these conditions: in finally suspending his belief, he seems to be acting on the basis of reasons. (This is even clearer in the Gallegos case: in finally managing to raise his prosthetic arm, Gallegos is plainly acting on the basis of reasons.) Plausibly, then, one can ϕ on the basis of R even if one does not occupy the agential position on ϕ -ing relative to R .

We have seen one feature of critical self-reflection that distinguishes it from otherwise similar cognitive processes, namely, the *directness* with which the thinker acts on her reasons. To capture this aspect of critical self-reflection, I've proposed that one engaged in such self-reflection occupies the agential position on preserving or revising attitudes. But occupying (and exploiting) the agential position is not sufficient for engaging in critical self-reflection, in Burge's sense. For revising an attitude directly on the basis of reasons need not be critical self-reflection.

Suppose that someone, inspired perhaps by reading William James, denies that ordinary norms of reasoning are sacrosanct, and maintains that some beliefs may legitimately be governed by practical norms. She decides that, within a certain sphere, she ought to accept thoughts she finds comforting, and reject those that upset her. After extensive practice, regulating beliefs according to this *wishful thinking* principle becomes second nature to her. So she occupies the agential position on regulating beliefs, relative to these emotional factors.¹⁰ But this way of regulating beliefs is not critical self-reflection in the relevant sense. To exclude such cases, we should reserve the term "critical self-reflection" for the preservation or revision (etc.) of an attitude directly on the basis of *appropriate* reasons, reasons corresponding to rational norms such as *beliefs should conform to evidence*.

On my proposal, critical self-reflection is the preservation or revision of attitudes directly on the basis of appropriate reasons. How does this proposal respond to the agentialist challenge to empiricism formulated in Section 2.3 above? Recall the key premise of Burge's argument:

¹⁰ Some will contend that the scenario described here is impossible, on the grounds that possessing the concept of belief requires grasping that belief aims at truth. The scenario could be redescribed as the decision to accept or regard as true those thoughts that provide comfort. Whether it is possible to accept a thought as true (or false) in *direct* response to its emotional tenor is, I think, a psychological question. I suspect that this is possible, but its possibility is not required by my argument. If it is impossible, then anyone who *actually* occupies the agential position on believing will occupy that position relative to (what she reasonably regards as) appropriate reasons.

- (3*) One cannot reason critically about one's attitudes unless one can know one's attitudes *non-empirically*.

Burge supports this premise as follows. In empirical self-knowledge, there are only “brute, contingent, non-rational relations” between the reviewing perspective, from which attitudes are known and evaluated, and the reviewed perspective, which includes the attitudes themselves (Burge 1996, 105). But critical reasoning about the attitudes—critical self-reflection—requires an “immediate rationally necessary connection” between these perspectives (*ibid.*, 109).

My proposal is, in effect, that a relation Burge would describe as “brute” and “contingent”—for convenience, we can think of this as a causal relation—can *constitute* a rational relation. When the realization that a belief is poorly justified directly causes the suspension of that belief, the relation between that realization (part of the reviewing perspective) and the belief (part of the reviewed perspective) is a rational relation. The relation between that realization and the belief's suspension is rational in that it satisfies the rational norm *beliefs should conform to evidence*.

Is this connection between the reviewing and reviewed perspectives an “immediate rationally necessary connection”, as Burge requires for critical self-reflection? It is “immediate” in that the realization that the belief is poorly justified unseats that belief directly. By contrast, for the detached juror the link between this realization and the belief suspension is indirect, as it is mediated by the exercise of imagination. And this connection is “rationally necessary” in the sense that the contents of these perspectives jointly contribute to what is rationally required of the thinker: given that realization, maintaining the belief would violate rational norms.

Importantly, however, in another sense this relation is contingent. The belief might withstand the realization that it is poorly justified. Of course, when a belief withstands the force of reasons, the process does not qualify as critical self-reflection. (This occurs not only in the detached juror case but also in the wishful thinking case: the wishful thinker violates rational norms when she conforms her beliefs to what she finds comforting, at least when what is comforting conflicts with her evidence.¹¹) But as the prevalence of belief perseverance shows, our deliberations often fall short of critical self-reflection.

On the plausible assumption that one counts as a rational agent only insofar as one can form or revise attitudes directly, on the basis of appropriate reasons, the proposal sketched here implies that:

- Necessarily, rational agents are capable of critical self-reflection.
- Necessarily, reasoning critically about a belief requires occupying the agential position on revising that belief, relative to one's evidence.

In effect, my proposal construes these statements as expressing *de dicto* necessities. This tactic echoes my earlier (Section 2) strategy for reconciling empiricism with the

¹¹ The wishful thinker may also violate rational norms in forming a belief for which she has strong evidence, if she forms it not on the basis of that evidence but only because it comforts her.

claim that, necessarily, rational creatures are capable of self-knowledge. While empiricism about self-knowledge is not incompatible with the statements above read as expressing *de re* necessities (see note 7), my construal of those statements is especially amenable to a broader empiricist outlook.

Some will object that this proposal neglects what is truly valuable about critical self-reflection, in contrast to simply registering the presence of an attitude, and evaluating it detachedly. But I think this objection is misplaced. On my proposal, critical self-reflection has two features that together distinguish it from allegedly “brute” operations on one’s own attitudes: its directness and its guidance by rational norms. These two features seem to me adequate to explain any special value attaching to critical self-reflection.

There is clearly instrumental value in being able to put one’s reasons into action *directly*: this is why Gallegos works so diligently to master his new prosthetic. And satisfying norms such as *beliefs should conform to evidence* may also be instrumentally valuable, assuming that evidence is appropriately linked with probability of truth and true beliefs possess instrumental value. Some believe that critical self-reflection has a kind of special non-instrumental value. I’m skeptical of this idea.¹² But my proposal can accommodate it. The claim that critical self-reflection is non-instrumentally valuable likely derives from the idea that there is intrinsic value to satisfying rational norms. The likelihood of satisfying rational norms is diminished in cases where reasons do not shape attitudes directly (as in the detached juror case). And it is further diminished when the thinker defies rational norms (as in the wishful thinking case). Thus, my proposal can make sense of the claim that critical self-reflection has greater instrumental and intrinsic value than other ways of shaping one’s attitudes.

5. Avowals

The notion of the agential position helps to elucidate both Moran’s notion of the “deliberative stance” and the phenomenon of avowals.

Moran imagines a situation involving Sartre’s akratic gambler. The gambler has concluded that he should quit gambling, but then engages in “empirical psychological reflection on himself” in an effort to predict whether he *will* quit.

What is wrong with the “direction of gaze” here, the shift to the theoretical or empirical perspective, is that it suggests that his reflection on his best reasons for belief or action still leave it an open question what he will actually end up believing or doing. This is not a stable position one can occupy and still conceive of oneself as a practical and theoretical deliberator. (Moran 2001, 94)

¹² Insofar as we deserve credit for any of our cognitive processes, the detached juror who works to overcome his persevering belief arguably deserves as much credit as the deliberative juror. The case of Corporal Gallegos demonstrates even more clearly that one does not deserve special credit merely for occupying the agential position.

The gambler endorses the reasons to quit gambling, as the reasons *to* act on: this is the sense in which they are his “best reasons”. These reasons compete with motivations (to gamble) that the gambler does not endorse. Moran contrasts the “empirical perspective” with the “deliberative stance”, the stance from which one decides which reasons *to* act on. To take the deliberative stance is, I suggest, to consider one’s attitudes with the presumption that one occupies the agential position on them, relative to those reasons one endorses as the reasons *to* act on. Adopting this stance, the gambler considers whether to quit gambling—in effect, whether to intend to quit—with the presumption that he can maintain this intention on the basis of the reasons he endorses. With this presumption in place, identifying the “best reasons for belief or action” does not “leave it an open question” how one will believe or act.

This brings us to avowals. On Moran’s view, one can *avow* only those attitudes that one takes to be transparent to one’s reasons. Plainly, one will not take an attitude to be transparent to (what one endorses as) one’s best reasons unless one presumes that one can directly do what one’s best reason dictates, regarding that attitude. So one can avow an attitude only if one presumes that one occupies the agential position on preserving or revising the attitude, relative to the reasons one endorses. This fits Moran’s characterization of the deliberative stance as “the stance of avowal” (*ibid.* 89). When the gambler utters “I will quit gambling” as an avowal rather than a mere prediction, he conveys that the intention to quit is based on reasons he endorses as the reasons to act on, and that he occupies the agential position on preserving this intention, relative those reasons. This is why avowals express commitments for which the avower is accountable. If the gambler continues to gamble, he is an apt target of criticism for not doing what is, by his own lights, both the thing to be done and within his powers.

Avowing an attitude conveys that the attitude reflects the agent’s best reasons. Does it also convey that these are *appropriate* reasons? This may depend on the attitude in question. Given the link that Moran envisions between belief avowals and transparency to evidence, belief avowals may include the presumption that the agent’s reasons for belief consist in her evidence, and hence are appropriate. In the case of intentions, the agent’s “best reasons” are those that reflect the values she endorses. What would count as an appropriate reason for *intending*—and indeed, whether there is some objective standard of appropriateness here, as Moran claims there is for reasons for belief—is less clear.

6. Empirical self-knowledge

The agentialist argues that avowals and critical self-reflection occur only when the thinker occupies the same “perspective” as the attitudes she avows or about which she reasons. Unsurprisingly, the agentialist construes this “same perspective” in agential terms. I’ve proposed that this agential dimension of avowals and critical self-reflection consists in the thinker’s occupying the agential position on the attitudes, relative to appropriate reasons: she can bring those reasons to bear on the attitudes directly.

The agentialist contends that empiricism about self-knowledge cannot do justice to the sense in which avowals and critical self-reflection occur from the same perspective as the target attitudes, because empiricism confines self-reflection and self-knowledge to a perspective that is distinct from the perspective of the attitudes known. To answer this

charge, I will argue that one may occupy the agential position on preserving or revising an attitude, relative to appropriate reasons, even if one's knowledge of that attitude is purely empirical. There are two kinds of self-knowledge to address.

- Self-knowledge of the attitude about which one reasons critically—e.g., the deliberative juror's knowledge that she believes that the defendant is guilty.
- Avowable self-knowledge about the outcome of one's critical self-reflection—e.g., the deliberative juror's knowledge that she (now) doesn't believe that the defendant is guilty.

An empiricist construal of each of these types of self-knowledge is compatible with my proposal about the agential dimension of critical self-reflection and avowals. We begin with knowledge of the belief that the defendant is guilty. The key point to notice here is that the detached juror may be on a par with the deliberative juror, as regards the epistemic basis for each juror's knowledge that s/he believes that the defendant is guilty. The crucial distinction between the two jurors is that only the deliberative juror occupies the agential position on revising that belief. This difference does not concern how the attitude is *known*, but rather how the thinker's reasons *affect* it.

Let's suppose, for illustration's sake, that this knowledge is based on inference from introspective evidence. Each juror considers what s/he believes on this matter by conjuring a mental image of the defendant; noticing that the image has an accompanying "guilty" phenomenology; and inferring that he believes that the defendant is guilty.¹³ (What's important here is not the details, but only that each juror's self-knowledge of the belief is based in empirical justification or warrant, and does not depend on any entitlement derived from rational agency.) Each juror then reflects on the prosecution's case and realizes that the belief is based on weak evidence. This reflection generates, for each juror, a reason to suspend the belief.

It is only at this point that the jurors diverge. The deliberative juror's reason to suspend the belief brings about her belief's suspension directly, whereas the detached juror's reason brings about his belief's suspension only indirectly (via his imagining the defendant elsewhere). Given my proposal, we can say (with the agentialist) that only the deliberative juror engages in critical self-reflection. On my analysis, this is because only the deliberative juror occupies the agential position on revising that belief, relative to appropriate reasons. But this difference emerges only after each juror recognizes that s/he has the belief in question. It does not imply that there is any difference between the jurors, as to how self-knowledge of the initial belief is achieved.

This brings us to the second kind of self-knowledge, concerning the outcome of critical self-reflection. This knowledge is arguably more salient for the agentialist's case, since it is only the outcome of critical self-reflection that is avowable.

¹³ Perhaps there is an alternative suspect, and each juror notices that his image of this other suspect is accompanied by an "innocent" phenomenology. Or perhaps each juror imagines the defendant committing the crime, and notices that this mental image has a "plausible" phenomenology about it (e.g. it seems particularly vivid or natural).

For the deliberative juror, the recognition that *my evidence for the defendant's guilt is weak* is closely linked with the belief *I do not believe that the defendant is guilty*. What is this link? On Moran's view, the epistemic right to believe that there is a link between one's evidence regarding X, and one's belief about X, is based in the "Transcendental assumption of Rational Thought": the assumption "that what I actually believe about X can be determined, made true by, my reflection on X itself" (Moran 2003, 406). But no transcendental reasoning is required to explain how the close link between the deliberative juror's belief and (what she regards as) her evidence puts her at an epistemic advantage, vis-à-vis self-knowledge, relative to the detached juror. For example, the empiricist can say that the deliberative juror *infers* that she doesn't believe that the defendant is guilty, from reflection on the weakness of her evidence. This inference is truth-preserving precisely because she occupies the agential position on the suspension of that belief: she can suspend it directly on the basis of her reason to do so. Alternatively, the empiricist could posit a monitoring mechanism, which takes as input one's (assessment of the) reasons for believing or intending, and delivers as output a belief about what one believes or intends. Such a mechanism will be reliable if the thinker occupies an agential position on regulating her attitudes, relative to her reasons.

The empiricist alternatives just mentioned are epistemically externalist. The empiricist also has internalist options for explaining the deliberative juror's superior epistemic position. The deliberative juror may have evidence to the effect that, at least when it comes to issues in which she has little emotional stake, her beliefs tend to conform to her evidence. (Her evidence for this conformance could be empirical: e.g., it could derive from introspection.) The detached juror would presumably have weaker evidence for the conformance between his beliefs and his evidence. In that case, the deliberative juror would be more strongly justified in inferring, from the fact that her evidence regarding the defendant's guilt is weak, that she does not believe that the defendant is guilty.

Moreover, the empiricist can explain the appeal of the transparency method, as a way of determining what one believes or intends. This method is appealing because we generally believe that our attitudes are sensitive to our reasons. If we didn't believe this, we wouldn't engage in practical and theoretical deliberations as a way of shaping our intentions and beliefs. Of course, this belief is sometimes false: attitudes sometimes resist the force of reasons. But Moran is surely correct that we generally *make* what he calls the Transcendental assumption of Rational Thought. Depending on the epistemological views she favors, the empiricist could claim that this assumption is justified or warranted. Or she could maintain that it is an ungrounded assumption, but that use of the transparency method yields self-knowledge nevertheless, because that method is sufficiently reliable or the beliefs it generates are "safe".

It may seem that these empiricist alternatives miss the point. They posit differences in epistemic resources, whereas what is fundamentally special about critical self-reflection, avowals, and the deliberative stance seems to be agential rather than epistemic. This latter idea may well be correct. As I said above, the fundamental difference between the two jurors seems to lie not in how they know their attitudes but, rather, in how each juror's reasons *affect* their attitudes. I have proposed ways of understanding this

difference in how reasons affect attitudes that are amenable to empiricism about self-knowledge. If empiricists can explain the relevant agential phenomena, they can answer the agentialist challenge. That challenge is further diminished by the fact that empiricists can explain how occupying the agential position could improve the prospects for self-knowledge.

7. Conclusion

Agentialism's defining thesis is that, in some cases, knowledge of one's own attitudes rests on a kind of *entitlement* that derives from rational agency and is irreducible to the kinds of epistemic factors countenanced by empiricism. This thesis entails that empiricism is false. I have not attempted to refute agentialism. I have argued only that a key component of the case for agentialism, namely the challenge to empiricism, fails.

My proposal is that the empiricist can do justice to the phenomena motivating agentialism as follows. Construe critical self-reflection as a cognitive process in which attitudes are directly sensitive to appropriate reasons, where this sensitivity may be purely causal-dispositional. A reason's directly affecting an attitude constitutes mental agency, in at least a minimal sense: it is the thinker's doing something on the basis of reasons. One engaged in critical self-reflection occupies the agential position on revising their attitudes. If we reserve the term "rational thinkers" for those capable of critical self-reflection then, necessarily, rational thinkers (sometimes) occupy the agential position on revising (some of) their attitudes, relative to appropriate reasons. To take the deliberative stance, in theoretical or practical reasoning, is to consider an attitude with the presumption that one occupies the agential position on believing or intending (etc.), relative to appropriate (rational or endorsed) reasons. And to avow a belief or intention is to convey that the attitude represents an exercise of agency: that, in believing or intending, one is directly doing what one's best reasons dictate.

Agentialists will reject this way of understanding critical self-reflection and avowals. They may think that my proposal neglects the special normativity inherent in practical and theoretical reasoning; overlooks the unique kind of agency involved in rational thought; ignores the fact that we must have some grasp of ourselves *as* rational agents, in order to coherently embark on the project of critical self-reflection; or discounts the significance of rational agency to our identities. At least some of these objections express agentialist commitments that partly constitute the rejection of empiricism. My task here was not to do justice to *all* agentialist commitments, but only to show that empiricism can respect the more plausible claims agentialists marshal in support of their view. In particular, empiricism can make sense of the idea that rational agency, relative to one's attitudes, is central to the phenomena of critical self-reflection and avowals.

For all I have said, agentialism may be true. But empiricism about self-knowledge withstands the agentialist challenge.¹⁴

¹⁴ Ancestors of this paper were presented at Central European University, the Virginia Philosophical Association, Rice University, the University of Graz, the University of Western Ontario, the SSPP, the University of Miami, the University of Alabama-Tuscaloosa, and the Rutgers Epistemology

Works cited

- Anscombe, G. E. M., 1963, *Intention*, second edition, Oxford: Blackwell.
- Armstrong, D., 1968/1993, *A Materialist Theory of the Mind*. London: Routledge.
- Bilgrami, A., 2006, *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press.
- Boyle, M., 2009, "Two Kinds of Self-Knowledge," *Philosophy and Phenomenological Research* 78: 133-164.
- Burge, T., 1996, "Our Entitlement to Self-Knowledge," *Proceedings of the Aristotelian Society* 96: 91-116.
- Burge, T., 2003, "Perceptual Entitlement," *Philosophy and Phenomenological Research* 67: 503-548.
- Byrne, A., 2011, "Transparency, Belief, Intention," *Aristotelian Society Supplementary Volume* 85: 201-221.
- Dao, James, 2012, "Learning to Accept, and Master, a \$110,000 Mechanical Arm". *New York Times* November 27, 2012.
- Dretske, F., 1994, "Introspection," *Proceedings of the Aristotelian Society* 94: 263–278
- Evans, G., 1982, *The Varieties of Reference*, Oxford: Oxford University Press (ed. J. McDowell).
- Fernández, J., 2013, *Transparent Minds: A Study of Self-Knowledge*. Oxford: Oxford University Press.
- Gallois, A., 1996, *The Mind Within, The World Without*. Cambridge: Cambridge University Press.
- Gertler, B., 2011, *Self-Knowledge*. Routledge.
- Kind, A., 2003, "Shoemaker, Self-Blindness, and Moore's Paradox," *Philosophical Quarterly* 53: 39–48.
- Korsgaard, C.M., 2009, "The Activity of Reason," *Proceedings and Addresses of the American Philosophical Association* 83: 23-43.
- Moran, R., 2001, *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton, NJ: Princeton University Press.
- Moran, R., 2003, "Responses to O'Brien and Shoemaker," *European Journal of Philosophy* 11: 402–19.
- Shoemaker, S., 1994, "Self-Knowledge and 'Inner Sense'," *Philosophy and Phenomenological Research* 54: 249–314.

Conference. Discussion at these events was enormously valuable. I'm especially indebted to Imogen Dickie, Harold Langsam, Sarah Paul, Jim Pryor, and Russ Shafer-Landau.