***Note to reader**: This is the penultimate draft of a paper published in Philosophical Studies under the same title. Please cite the published version of the paper. The final publication is available at Springer via* http://dx.doi.org/10.1007/s11098-016-0685-4.

Embedded Mental Action in Self-Attribution of Belief

Abstract You can come to know that you believe that *p* partly by reflecting on whether *p* and then judging that *p*. Call this procedure "the transparency method for belief." How exactly does the transparency method generate *known* self-attributions of belief? To answer that question, we cannot interpret the transparency method as involving a *transition* between the contents *p* and *I believe that p*. It is hard to see how some such transition could be warranted. Instead, in this context, one mental action is *both* a judgment that *p* and a self-attribution of a belief that *p*. The notion of *embedded mental action* is introduced here to explain how this can be so and to provide a full epistemic explanation of the transparency method. That explanation makes sense of first-person authority and immediacy in transparent self-knowledge. In generalized form, it gives sufficient conditions on an attitude's being known transparently.

Keywords Self-Knowledge · Belief · Transparency · Mental Action · Agency

One way to know that you believe that *p* involves coming to judge that *p*.[1] As Evans (1982) famously observed, in answering the question "'Do you think there is going to be a third world war?', I must attend … to precisely the same outward phenomena as I would attend to if I were answering the question, 'Will there be a third world war?'" (p.225).[2] Call this way of answering this question "the transparency method for belief," and the attributions it generates "transparent self-attributions of belief." What explains the *knowledge* status of such transparent self-attributions?

There is a challenge facing any such attempt at epistemic explanation: the transparency method seems to involve a *move* from a judgment about the world to a judgment about one's own beliefs. But the content of the former judgment does not stand in any ordinary justificatory relationship to the latter. The task at hand is to describe the connection between a judgment that *p* and a self-attribution of a belief that *p* that makes sense of the epistemic warrant involved in the transparency method.[3]

---

[1] I'd like to thank an anonymous reviewer for invaluable commentary on a previous draft.

[2] Evans's (1982) characterization of the transparency method is the most often cited, but—as Moran (2001) notes—Roy Edgley (1969) seems to have provided the first characterization.

[3] Here I follow Burge (1996) in taking "warrant" to be a broad term that covers epistemic entitlement, inferential justification, evidentiary support, and so forth (see pp.93-94).

Ideally, a successful explanation will also make sense of first-person authority about belief—the authority you enjoy over all others in attributing beliefs to yourself. The transparency method is one strong candidate for a method of belief attribution that enjoys epistemic privilege in its first-personal use.

This paper offers a new epistemic explanation of the transparency method. Crucial to that explanation is the fact that the transparency method does not involve any *move* between *p* and *I believe that p*. This fact has been recognized in the past, but never fully explained: how can a judgment that *p also be* a self-attribution of a belief that *p*? This paper has two main goals: first, to explain how just one mental action can be both; and second, to use *that* explanation to explain why transparent self-attributions of belief have the status of knowledge.

1 Attempts to explain the epistemology of the transparency method

Why accept the claim that you can come to know that you believe *p* in part by considering whether *p*? Besides a certain phenomenological appeal, the suggestion has two clear attractions. First, it coheres elegantly with Moore's famous observation that assertions of the form "*p*, but I don't believe that *p*" are absurd.[4] If judging that *p* enjoys a close connection with knowing you believe that *p*, the second conjunct may directly conflict with knowledge gained by judging the first.[5] Additionally, accepting the transparency method as a way of knowing your beliefs has the potential to anchor a metaphysically economical explanation of first-person authority about belief.

In order to offer some such theory, though, we must first understand how the transparency method produces *known* self-attributions of belief. And now we face a prima facie explanatory challenge: a challenge to demonstrate how an apparent move from a judgment that *p* to a self-attribution of a belief that *p* is epistemically warranted.

Two types of attempt to explain the warrant involved in the transparency method have been particularly influential. One treats the move from *p* to *I believe that p* as a form of inference, and the relevant warrant as inferential warrant. The other takes the rational agency we enjoy with respect to our own beliefs to explain the relevant epistemic warrant. Both explanatory strategies face substantial objections. By coming to understand these objections, we can start to see the shape of a more promising epistemic explanation.

1.1 Gallois and Byrne on Moore inferences

According to Gallois (1996), using the transparency method involves making an *inference* from *p* to *I believe that p*—a "Moore inference".[6] Though Moore inferences are not deductively valid, inductively valid, or abductively plausible, Byrne (2005, 2011,

[4] See Moran (2001), Barnett (2015), Williams (2004), Gallois (1996), Byrne (2011), and Shoemaker (1994) for more on the connection between transparency and self-knowledge.

[5] There is naturally far more to say on this point to make proper sense of an explanation of Moorean absurdities. Note, too, that not all those philosophers aiming to explain the absurdity of Moorean assertions and judgments accept that an epistemic explanation is the right kind to give. See Green and Williams (2006) for arguments on both sides of this divide.

[6] Gallois (1996) names these "Moore inferences" because reasoning in this way explicitly avoids commitment to the Moorean absurdity "*p*, but I don't believe that *p*" (p.46).

2012) argues that they are "not without epistemic merit" (2011, p.206). Their inferential schema is "*strongly self-verifying*": self-verifying because "inference from a premiss entails belief in that premiss," and *strongly* self-verifying because the conclusion will be true even when the premise is false (p.206). The ensuing self-attributions are "*safe* in the sense that they could not easily have been false" (p.206-7).

Byrne recognizes that this point "falls short of a *demonstration* that reasoning [in this way] is knowledge-conducive": neither strong self-verification nor safety ensures warrant (p.207). However, he has not set out to prove that such reasoning *is* knowledge-conducive; this has not been called into question. What is at issue is how to explain privileged and peculiar access. To say that one has privileged access to one's own beliefs is just to say that "beliefs about one's mental states are more likely to amount to *knowledge* than one's corresponding beliefs about others' mental states" (p.202, emphasis added). To say that one has peculiar access to one's own beliefs is just to say that "one has a way of *knowing* about one's mental states that one cannot use to come to know about the mental states of others" (p.202, emphasis added).[7] To explain why we all generally enjoy privileged and peculiar access is also to explain first-person authority.[8]

Byrne claims to have given privileged and peculiar access a "satisfying explanation" (p.207). But these notions are explicitly epistemic. Why, then, accept a view that fails to explain how the transparency method is knowledge-conducive?

Can the inferential interpretation of the transparency method at least help us along the way to some such properly epistemic explanation? It can do that only if Moore inferences share some non-trivial structure with inferences in general. In recent years, however, more and more disanalogies between Moore inferences and inferences in general have been recognized by respondents to Byrne (Barnett 2015; Boyle 2011; Brueckner 1998; Moran 2012; Silins 2012).

Most obviously, Moore inferences have a generality problem. As Boghossian (2014) has pointed out, "something that seems very central to our inferential abilities [is that] they are both *general* and *productive*" (p.12). But Moore inferences cannot be run in the past or future tense (Byrne 2011, pp.204-206). They fail in the third person. And Moore inferences are absurd in hypothetical reasoning (Barnett 2015; Brueckner 1998).

Inferential conclusions also inherit the strength or weakness of the justification for their premises (Barnett 2015; Setiya 2011; Silins 2012; Shoemaker 2009).[9] Moore inferences' premises, on the other hand, are meant to justify their conclusions irrespective of the justification—or even truth—of their premises.[10]

It is crucial that inferences to judged conclusions depend on the truth of their premises' *contents*. "A transition … counts as inference only if the thinker takes his

[7] Byrne (2012) credits Gilbert Ryle (1949) with coining the phrase "privileged access."

[8] Some, however, reject the converse. Davidson (1984/2001, 1987/2001) and Wright (1989), among others, offer non-epistemic explanations of first-person authority.

[9] This is related to the distinct point that knowledge cannot be gained by way of false lemmas. Some take that to be the lesson of Gettier's (1963) famous thought experiment. Byrne (2011, p.207) correctly points out, however, that one need not draw *that* lesson from Gettier; instead, we might restrict knowledge to *safe* cases. Moore inferences are, as he points out, as safe as it gets.

[10] See Barnett (2015) for extensive careful discussion of the epistemological ramifications of accepting Moore inferences as such. He also points out two further disanalogies between Moore inferences and inferences in general which, for reasons of space, I have omitted here.

conclusion to be supported by the presumed *truth* of those other beliefs," writes Boghossian (2014, p.4, emphasis changed).[11] On the other hand, what actually grounds the truth of any transparent self-attribution of belief is the *attitudinal* aspect of the judgment performed in the transparency method: its being judged, rather than, say, merely entertained.

Due to these significant disanalogies, it becomes difficult to see how the Gallois-Byrne proposal could add to our epistemic understanding of the transparency method. It's not clear what is being borrowed from ordinary cases of inference that could help us make sense of the warrant for transparent self-attributions. The challenges to this view are also challenges for *any* view of the transparency method on which using that method involves a *transition*—inferential or otherwise—from *p* to *I believe that p*. How could this transition be warranted, if not by way of any recognizable inferential justification?

1.2 Moran on rational deliberation

Moran (2001, 2003, 2004) rejects the inferential view of the transparency method. He argues instead that your first-person *epistemic* authority on the matter of what you believe derives from your *deliberative* authority. You have deliberative authority insofar as you have "the capacity to play a constituting role in determining the psychological facts themselves"—that is, the capacity to make up your mind in a rational way (p.146). Moran (2001) implies that to understand the special first-personal "stance of rational agency" is already to understand the epistemology of the transparency method (p.149).

The suggestion has great intuitive appeal, but Moran's (2001) original presentation of it does not provide much further detail about the relevant epistemic explanation. In response to questions about how this epistemic explanation is meant to work (O'Brien 2003; Shoemaker 2003; Wilson 2004), Moran (2003, 2004) has clarified what he takes to be the structure of the warrant involved in the transparency method:

> if the person were *entitled* to assume, or in some way even obligated to assume, that his considerations for or against believing P (the outward-directed question) actually determined in this case what his belief concerning P actually is (the inward-directed question), then he would be *entitled* to answer the question concerning his believing P or not by consideration of the reasons in favor of P (Moran 2004, p.457, my emphasis).

The warrant at issue, then, is a form of entitlement to a guiding assumption that your deliberative reasons settle your beliefs.

On Moran's view, you can't avoid making this entitling assumption: it follows from the nature of deliberation that you take yourself to be settling your beliefs. And you have the relevant entitlement whenever the assumption is true—whenever deliberation does indeed determine your beliefs. However, you can also "lose the right" to the assumption when it is not true (2003, p.406). Moran emphasizes several times that your rational deliberations can, and often do, fail to determine your beliefs.

---

[11] Cf. Frege (1979, p.3), quoted by Boghossian (2014, p.4): "To make a judgment because we are cognizant of other truths as providing a justification for it is known as *inferring*."

This entitlement-based modification of the account is helpful. With it, Moran's view can make sense of how you know your transparent self-attributions in at least some cases. Because the entitlement is attached to a specially first-personal assumption, the account also offers some promise with regard to explaining first-person authority.

However, the account has trouble explaining first-person authority as a *general* phenomenon, precisely because Moran takes the assumption to be frequently false. That implies that your transparent self-attributions of belief are also frequently false. Not only will they be false: other people will also be better placed epistemically to *recognize* their falsity. From a third-personal standpoint—say, a standpoint of someone who knows you and your dispositions very well by way of close observation—it may be clear that your deliberating in this way does *not* settle what you believe. But *you* are always constrained to assume, inaccurately, that your deliberation *does* settle your beliefs. Most of the time, then, the necessary assumption about deliberation (to which you are *sometimes* entitled) will be an epistemic liability, rather than an asset.

The problem can be partially fixed by abandoning the independently plausible claim that the entitling assumption is often wrong. But even if the assumption were universally correct, the view would remain incomplete in another distinct way. Given Moran's presentation of the assumption, it seems like you would need to identify your considerations for or against believing that *p as such* in order to use the assumption to come to know whether you believe *p*. If that is indeed the case, then we need to understand how you could come to know what your considerations for and against belief are. Moran's explanation will remain incomplete until we have some further explanation of that sort of self-knowledge—or, at the very least, good reason to think that we have such self-knowledge in all the relevant cases of transparent self-attribution of belief.[12]

Moran might simply deny that you need self-awareness of your considerations to apply the relevant assumption to come to know what you believe. But if that is the case, it's difficult to see how making the assumption that your considerations for and against believing that *p* determine what you actually believe could help explain your knowledge of your own beliefs at all. What is needed here is a better understanding of the types of self-awareness involved in using the transparency method to self-attribute belief.

1.3 Towards a better explanation

As seen above, explaining the epistemic warrant for a move from *p* to *I believe that p* is prohibitively difficult. Why not, then, resist understanding the transparency method as involving any such *move* at all? Boyle (2009b, 2011), Crane (2001) and Setiya (2011) have all advanced positions of this type, which I will call "no-move views."

Each such view faces further explanatory challenges that have not yet been met. In particular, all three views posit the identification of types of attitudes or acts that seem entirely distinct. Yet none fully explains how to understand such identifications.

---

[12] It is worth noting that self-knowledge of reasons might be transparent in a way directly analogous to self-knowledge of belief: you may well know your own considerations for or against believing that *p* by considering what in fact provides evidence for (or logically implies that) *p*. Yet this observation cannot alone complete Moran's explanation. The same issues that motivate this paper—questions about *why* judging that *p* leads to knowing you believe that *p*—would then apply in analogous ways for knowledge of your considerations for and against belief that *p*.

Consider Boyle's (2009b, 2011) picture, the most developed of the three. Boyle (2011) argues that the transparency method merely manifests implicit knowledge. Writes Boyle: "The step [involved in the transparency method], in other words, will not be an inferential transition between *contents*, but a coming to explicit acknowledgment of a *condition* of which one is already tacitly aware" (2011, p.227). Crucially, what is made explicit here are not *two* states—a belief, and knowledge of that belief—but just one: "believing *P* and knowing oneself to believe *P* … are two aspects of *one* cognitive state—the state, as we might put it, of knowingly believing *P*" (p.228).

Partly to explain this puzzling claim, Boyle (2011) echoes Moran's point that our rational authority over our own beliefs must play a role in explaining how we know about them: "we have privileged self-knowledge of these sorts of attitudes *because* they are expressions of our capacity for rational self-determination" (p.237, emphasis added).[13] As with Moran's (2001) discussion, though, this point alone cannot provide the epistemological understanding we are after. And if it is the identification of belief and knowledge of belief that is meant to provide such epistemological insight, we need to better understand *how* two such apparently distinct states could be one and the same.[14]

The denial of any *move* in the transparency method also appears in a proposal by Crane (2001), who claims that "being conscious that one believes that *p* need involve no more than being conscious that *p*" (p.108), and a passage by Setiya (2011), who writes that the thought that *p* and the thought *I believe that p* "need not even be distinct from one another" in the context of the transparency method (p.187).[15]

Here Crane and Setiya each disagree with Boyle on one key point. While Boyle (2011) argues that belief and *knowledge* of belief are one and the same state, Crane claims that *consciousness* that *p* can also be consciousness that one believes that *p*, and Setiya suggests that the *thought* that *p* and the thought that one believes that *p* can be identical. Yet each of these proposed identifications is just as puzzling as the last.

To move towards a compelling no-move view, we need to get clear on three points. First: what should be identified with what in the transparency method? Second: when and how are these two apparently distinct attitudes or acts one and the same? And finally: can this identification help us understand the epistemology of the transparency method?

---

[13] Here (2011) and elsewhere (especially 2009a) Boyle makes plausible that a belief is "an exercise of agency" (2009a *passim*) insofar as its subject controls and guides its development.

[14] This challenge is sharpened by the fact that subjects can apparently fail to know their own beliefs. Boyle (2011) is constrained to deny this. He writes: "when a belief is present but not consciously accessible, *so too is the knowledge of that belief*" (p.229). This is a deeply counterintuitive conclusion. It certainly seems to be the case that you can discover, with surprise, that you have some belief that you previously never knew you had—not even tacitly.

[15] Compare Heal (2002) and O'Brien (2005). Heal argues that the judgment that *I believe that p* as made in the course of the transparency method is also a judgment that *p*, because it shares the "long-term consequences" of the judgment that *p* (p.17). Yet it is not clear why meeting this condition requires identification of these judgments. O'Brien may also have a no-move view about self-knowledge of *judgment*, if not belief. She writes: "concluding that '*P* is true', on considering whether *P* is true, is … equivalent to the subject realising the practically known possibility of judging that *P*" (p.594). This practical knowledge just is the knowledge that one is judging that *P*. However, O'Brien explicitly avoids extending this account to belief (pp.599-600).

2 Mental action and embedded mental action

To answer those questions, I'll introduce the notion of *embedded mental action*. To do that, I'll begin with a discussion of intentional mental action more generally.[16]

2.1 Intentional mental actions and action awareness

Mental actions are things one does mentally. Imagining what your wedding will be like can be a mental action. Recalling what your doctor told you can be a mental action. Supposing that $x = 4$ can be a mental action.

Three facts about mental action will be crucial to this discussion. First, mental actions can be *intentional*. Second, intentional mental actions have conscious contents. Third, in performing an intentional mental action, you have contrastive awareness of the kind of thing you are doing in thought.

Importantly, we do *not* need to settle any other controversies here about mental actions. We do not need to settle, for instance, whether mental actions are *necessarily* or *always* intentional. Nor do we need to commit to a view about what it is that makes a mental action intentional.[17] I'll set aside these questions for now.

To see how mental actions can be intentional, consider an example. You can intentionally imagine, say, what your wedding will be like. That is not to say that *imagining* is intentional by definition, but just that you *can* imagine something intentionally. Nor is it to say that all aspects of your imagining are under your direct control. It might just happen to occur to you that it might rain on your wedding day, bringing to mind visions of soaked guests and stained shoes. That is compatible with your ongoing task of imagination being intentional.

Compare recollection. You can involuntarily recall what your doctor said, or you can intentionally recall it. When done intentionally, recalling something can require effort and sharp focus. It requires trying to be guided only by what actually happened, but that does not mean you have direct control over what you ultimately recall. Indeed, it wouldn't properly be *recalling* anything if you got to choose anything to bring to mind. Your act of recollection's being intentional is not only compatible with, but also implies, your not having arbitrary control over what comes to mind in performing your task.

The fact that mental actions can be intentional is crucial for the purposes of this paper. For that reason, it is worth stressing how strange it would be to deny that mental actions can be intentional. Someone who could never perform any mental actions *intentionally* could never set her mind to some particular mental task and then do it—even a task as simple as adding twelve and nineteen, or deciding where to go for lunch.

---

[16] For the following characterization of mental action I am indebted quite generally to O'Brien (2007), O'Brien and Soteriou (2009), Peacocke (2008), and Ryle (1971a-c).

[17] There is one delicacy here. Below, I suggest that deciding what to do (which implies forming an intention) can be intentional. If so, and if something's being intentional is partly a matter of having an intention to do it, then not *all* intention formation can be intentional, on pain of vicious regress. Sometimes you must be able to decide what to do without doing *that* (so deciding) intentionally. An anonymous reviewer drew my attention to this subtle point.

She might *in fact* add sometimes, and might *in fact* decide on restaurants, but she would be entirely passive with respect to her train of mental events. She might sometimes be able to follow or even participate in a conversation, if her thoughts cooperated nicely enough. But she could never *intentionally do* anything requested of her in simple communication. She could never intentionally think of a playing card for a magic trick, or intentionally consider whether the soup would taste better with more salt. The picture that emerges is beyond belief. Someone who could never perform intentional mental actions would be absurdly cut off from directing her own mental life.

When mental actions are intentional, they have conscious contents.[18] To say that you are trying to imagine what your wedding will be like is also to say that your conscious stream of thought has a certain kind of content (perhaps involving a banquet or a rabbi). Yet it is not *only* when imagination is active and intentional in this way that its contents are conscious. Involuntary imagined visions can pop into your mind as well.

The fact that mental actions can be intentional has another crucial implication: since you can do things intentionally in thought, you can be aware of what you are doing in thought. In other words, intentional mental actions involve *mental action awareness*.

This claim is a particular application of Anscombe's (1957) famous claim that doing something intentionally involves knowing what you are doing under a particular description—the description under which you intended to do it.[19] The same points that made this claim plausible for action in general also apply to mental action in particular.

As Anscombe noted, intentional actions are actions to which a certain kind of "Why?" question has application. That "Why?" question is a request for reasons that recommended a particular action to you. One important way of rejecting the applicability of that "Why?" question is to say that you did not know that you were doing that thing (under that description) at all. But if not knowing you are doing something (under a particular description) implies that the "Why?" question does not apply, and all intentional actions are ones to which the question *does* apply, then lacking awareness of what you are doing implies that you are not doing it intentionally. In contraposition, then, doing something intentionally involves knowing that you are doing it.

Consider these points as applied to *mental* action. When you are doing something intentionally in thought, you can be asked why you are doing it in the same special sense. For example: if you are trying to plot the social ruin of your rival, I can ask you why, and you might tell me that you need to be prepared in case the time comes for revenge. Similarly, if you are intentionally assessing your daughter's skill in tennis, I can ask you why, and you can tell me that you need to decide which lessons to book for her.

---

[18] In personal communication, [NAME REDACTED FOR ANONYMITY] has suggested that there could be intentional mental actions with unconscious contents. Though I disagree, I leave this subtle controversy aside for now. Even if some intentional mental actions lack conscious contents, certainly *some* mental actions have conscious contents. That weaker claim is all that is strictly required for the purposes of the epistemic explanation provided in Section 3 below.

[19] See also Davidson (1963/2001) on the connection between intentional actions and particular intensional descriptions of those actions. For more on mental action awareness generally, see O'Brien (2007), O'Brien and Soteriou (2009), and Peacocke (2008).

In either case, you can reject the legitimacy of my question by telling me that you didn't know what you were doing.[20] You might question whether you were really *plotting* the social ruin of your rival, or just imagining a scenario while making no real decisions. If, in your passive train of thought, you did in fact come to a judgment that your daughter has no real skill at tennis, you might simply not know whether that was a real judgment or merely a case of entertaining a hypothetical. Crucially, the uncertainty in either case would attach to the attitudinal aspect of the thought, which is precisely that aspect over which you sometimes have intentional control. The contrast is between knowing that you're doing *this* sort of thing (judging) or doing *that* sort of thing (supposing), rather than a contrast between being aware or unaware of the contents of those same thoughts.

In the case of intentional mental action, the awareness in question may not be particularly conceptually rich. Indeed, it need be no richer than the description that characterizes your action as intentional. For example: to have action awareness of an act of imagining, you need not be able to think explicitly in terms of acts of imagining. Rather, you can have more basic *contrastive awareness* of what you are doing: you can be aware, that is, that you are doing *this* sort of thing (e.g. imagining one's wedding) rather than *that* sort of thing (e.g. recalling one's wedding). The contrast to which you are sensitive in having mental action awareness is a contrast of attitudinal stance between different kinds of mental actions, but you need not conceptualize the difference *as* a contrast in attitudinal stance.

Mental actions can indeed be intentional. When they are intentional, they have conscious contents. To perform an intentional mental action is also to have awareness of what you are doing—at least of a basic, contrastive kind.

2.2 Judgment as an intentional mental action

As with imagining, judgment is not a mental action by definition, but *some* judgments are intentional mental actions. Those judgments, like all intentional mental actions, involve action awareness and have conscious contents.

It is worth clarifying the sense in which judgment can be intentional. Judgment can be intentional insofar as you can set out to *judge* some things, rather than, say, imagine some things. That is what you do when you set out to determine what's true. You can also set out to make a judgment that matches some particular content criterion—e.g. a *judgment about topic T* or a *judgment whether p*.

We need not accept voluntarism about judgment to see that judgments can be intentional. Though judgment can be intentional, it is not possible to decide at will the *precise* content of one's judgments: you cannot, without regard for the truth of some proposition *p*, will yourself to judge that *p*.[21]

As Shah and Velleman (2005) have recognized, the falsity of voluntarism about judgment proceeds from the nature of judgment. Judgments are just those mental actions

[20] In order to make best sense of my asking you this question when you yourself do not report to me what you are doing, assume that I have insight into what you are currently doing in thought just by watching you: I know that particular scowl you make when looking at your rival's portrait is caused by your plotting, or I know that your crestfallen expression while watching your daughter's tennis match is a result of a real judgment that she has no future at Wimbledon.

[21] See Williams (1976) for a related argument against doxastic voluntarism.

"that [are] governed, both normatively and descriptively, by the standard of truth" (p.499). That claim involves two important observations. First, judgments are in fact regulated by truth-seeking considerations. For example, when I gain conclusive evidence that some *p* is false and recognize that evidence for what it is, I will (generally) not judge that *p*. Second, judgments are normatively assessed relative to the standard of truth.

2.3 Embedded mental action

There are (at least) two different ways of individuating mental actions. First, mental actions can be individuated by their thought-types: *imagining* that you won gold in the hundred-meter dash is different from *recalling* that. Second, mental actions of the same thought-type can be individuated by their thought-contents: judging that you won *gold* is a different action than judging that you won *silver*. For the purposes of this paper, take a *type* of mental action simpliciter to be a class of judgments as individuated in *both* ways. An example of a type of mental action is *judging that you've won gold*.[22]

Here's an abstract definition of embedded mental action (to be illustrated below with examples): an embedded mental action is any intentional mental action of some type *T* that also belongs to another type *U* in virtue of your having antecedently conceptualized its content in some particular way (*de dicto*) in your intention to judge. In the context of an ongoing purposeful mental task, when you think of the contents of some of your upcoming thoughts (perhaps just those meeting a description *d*) as being *F*, in performing each such mental action (that meets *d*) you *already* take that particular content to be *F*. No further move is required for you to take those contents to be *F*.[23]

Two concrete examples will help clarify what an embedded mental action is.

First: imagine you are writing a short piece of fiction about acclaimed actors Lupita Nyong'o and Idris Elba. In this context, you need to *think of something Lupita could say to Idris* in response to a compliment. To do that, you call to mind various sentences, e.g. "You flatter me!" In imagining each of these sentences, you are not *just* thinking of a sentence; you are *thinking of something that Lupita could say to Idris*.

It is your antecedent conceptualization of such sentences as things Lupita might say to Idris that *makes* each of these mental acts more than just an act of thinking of a sentence. For what else could make each such action into an act of thinking of something that Lupita could say to Idris? Perhaps you could conjure a visual image of a person matching Lupita's description saying each sentence to someone matching Idris's description. But no such visual image is *required* to imagine Lupita saying something to Idris; someone with no capacity for visual imagination could do that. Nor is it sufficient:

[22] Broader classes of mental actions I will call "kinds": thus judging is a *kind* of mental action.

[23] This phenomenon is not particularly circumscribed to *mental* actions rather than actions in general. Compare the following case in non-mental action (with thanks to John Campbell): conceptualizing what you're about to do as *aiming for target one* partly makes it the case that what you go on to do, when you let fly, is *attempt to hit target one*. I focus here on mental action for simplicity and brevity, but it's important to note that this point about embedded mental action proceeds from a more general point about action and how it is conceptualized in thought.

the visual image itself wouldn't settle whether that was *really* Lupita saying something to Idris in your imagination.[24]

It is just your antecedent conceptualization of such sentences *as* things that Lupita might say to Idris that makes it the case that each of these embedded mental actions is not only an act of imagining a sentence but also an act of thinking of something Lupita could say to Idris.

Consider, now, a different example. In the example of embedded mental action just considered, you were not constrained to call to mind new or original sentences in order to complete your mental task. You could instead find things for Lupita to say by recalling some things that people have said to you, for instance. Any mental action that involves entertaining words in thought will do for the task at hand.

But not all embedded mental action works that way. Other mental tasks demand particular *kinds* of embedded mental action, e.g. judgment. Consider a second example: the task of *thinking of something you've learned about Lincoln*. In this case, it's not enough to think what *might* be true of Lincoln; you need to get at what *is* true about Lincoln, as learning is factive. You must restrict yourself to mental actions that are regulated for truth: judgments.

Moreover, the task is not just to find some fact about Lincoln. It's to think of something *you have learned* about Lincoln. You must judge, of some *p* pertaining to Lincoln, that you've learned that *p*. That judgment involves committing to the truth of *p*. To make some such judgment about what you've *learned* about Lincoln, you must restrict yourself to making judgments about what's *true* of Lincoln.

As a competent speaker of English with the concept LEARN, you recognize that you must make judgments for this task (even if you don't think of doing that *as* making judgments).[25] In other words, you understand, at least implicitly, that *this* sort of mental action (judgment) is appropriate for finding out what you have learned. This understanding contributes to your intention to *judge* rather than, say, imagine.

In thinking of something you've learned about Lincoln, all along you are thinking of the contents of your judgments *as* things that you've learned about Lincoln. For that reason, you *already recognize* the content of any such judgment as something you've learned about Lincoln just in making that judgment. You don't need to check whether you've just made a judgment, since you are aware of what you are doing. There are no two distinct actions here—making a judgment about Lincoln and making a judgment about what you've learned about him. You do the latter just in doing the former. In this case, the *embedded mental action* just is that first-order judgment about Lincoln.

2.4 The transparency method

How can embedded mental understanding help us understand the transparency method?

In my view, you use the transparency method just when you do all of the following in the course of a single extended mental task.[26] First, you intentionally set out to self-

[24] Conceptualizing these people in the right way would do the trick, but that is precisely the point being made here.

[25] In this paper I follow the convention of using SMALL CAPITALS to refer to concepts.

[26] In fact, these conditions cannot *quite* be said to be sufficient for use of the transparency method, because they do not rule out deviant causal chains of the kind famously discussed by

attribute a belief meeting some content requirement (e.g. a belief about some topic, or a belief whether *p*). As a result, you intentionally set out to make a judgment with some content that meets the relevant requirement, already understanding the content of that judgment as the content of a belief you have (though you need not think of what you are doing explicitly in terms of judgment). As a result of *that*, you actually do make a judgment with a content that meets the requirement on the belief to be self-attributed—and *in so doing*, you judge, first-personally, that you have a belief with that content.

The judgment that meets the content requirement is the *embedded mental action* here. Your antecedent conceptualization of the content of this judgment as the content of a belief of yours makes it the case that you *already* self-attribute a belief with the same content just in making this judgment.

Just as in the case of thinking of something you've learned about Lincoln, here you must implicitly understand that judgments (perhaps not understood, by you, in that way) are the right sorts of mental actions to perform in order to get at your beliefs on the matter. Because you have the concept BELIEF, you recognize that you couldn't just imagine things for the same purpose. This time, that's not because belief is factive like learning; belief is not factive. Here, your recognition that you must use judgment (perhaps not thus conceptualized by you) stems from implicit conceptual understanding that the attitude you take in judging is the same sort of attitude you have as when you have a belief about the world. You want to end up with a judgment about what you believe, and you can't willfully deceive yourself into thinking that the contents of some mental actions that aren't judgments (e.g. imaginings) are also the contents of your beliefs. Thus you're constrained to make judgments to self-attribute a belief whether *p*.

3 The epistemic explanation

Does this interpretation of the transparency method—as involving embedded mental action—offer any hope for a better epistemic explanation of the method as a whole? The point of this section is to demonstrate that it does just by giving such an explanation.

The epistemic explanation has four parts. The first three parts are: an explanation of the *truth* of transparent self-attributions; an explanation of their *warrant*; and an explanation of how they come to be *believed*. The final part of the explanation demonstrates the sufficiency of the whole explanation by showing that the warrant in question cannot be disqualified in any of the sorts of ways discussed by Gettier (1963).

3.1 Truth

Transparent self-attributions are guaranteed to be true because judgment is sufficient for contemporaneous belief.[27] Actively recognizing the truth of *p*—that is, judging that *p*—must involve at the very least having a *momentary* belief that *p*. Why is that? Because belief just is the mental state "that is governed, both normatively and descriptively, by the

---

Davidson (1973/2001). Suffice it to say, for now, that the transparency method is used *only if* all the listed conditions are met *and* the intentions of the first two conditions cause the embedded mental action of the third and fourth conditions *in the right way*.

[27] I do not, however, endorse the claim that judgment at some time *t* is sufficient for belief at any *other* time *t'*, or for any *interval* of time *T*.

standard of truth," and judgment just is the mental action governed in precisely the same way (Shah and Velleman 2005, p.499).[28] For that reason, one cannot perform the action without being in the corresponding state. The ontological difference between action and state does not block the implication (although the difference *does* block the converse). Boyle (2009a, pp.11-14) has made this point in a particularly powerful way.

On scrutiny of putative examples of judgment without belief, these can be shown to involve judgments and beliefs with distinct temporal profiles. Silins (2012), for example, presents an example of an "accidental" judgment, a "performance error which fails to reflect an underlying belief … you 'blurted out' that *p*, either in speech or merely in thought, consciously endorsing the proposition that *p*, yet failing to have a standing belief that *p*" (p.308). Silins here relies on the distinction between what he calls "standing belief" or "underlying belief"—a belief state one is in for some more extended interval—and momentary belief.[29] A better way of understanding Silins's example is as an example of rapid doxastic change, or a case in which the agent doesn't really ever judge that *p*.

Peacocke (1998) offers another example: "someone may judge that undergraduate degrees from countries other than her own are of an equal standard to her own, and excellent reasons may be operative in her assertions to that effect. All the same, it may be quite clear, in decisions she makes on hiring, or in making recommendations, that she does not really have this belief at all" (p.90). Once again the possibility in question trades on different timescales for belief and judgment. Peacocke illustrates that you might be best described as *lacking* a particular belief over an extended period of time even though you make genuine judgments at moments during that interval with the corresponding content. Even if that is true, it would be strange to insist that *at no point* during this interval, not even during those moments of judgment, does one have the relevant belief. One cannot fail, at the moment of judgment, to have the corresponding belief, although considerations about what it is to have a belief over some extended interval might bring us to admit that one doesn't "really" have the belief during that time.[30]

Further resistance to the claim that judgment is sufficient for belief may derive from a metaphysical view of belief on which it is essentially a complex dispositional state

---

[28] I do not mean, here, simply to *interdefine* judgment and belief as (for example) Crane (2001) does: "judgment is the formation of belief" (p.104). I take judgment that *p* to be possible when one already has the belief that *p*. If judgment just is, by definition, the formation of belief, then either this would not be possible, or one would have to be able to supplant a pre-existing belief that *p* with a *new* belief that *p* just by judging that *p*. Both options seem unattractive.

[29] There is no in-principle limitation on how short-lived genuine beliefs can be. There is no absurdity in saying "I really believed that for just one moment." Consider the following example. In a hurry to catch a flight, I rush through airport security and pause, uncertain which gate is mine. I glance at my boarding pass and see "34B." I start towards gate 34B, before realizing, just one moment later, that "34B" is my seat and my gate is instead 11B. I pivot on my heel and take off in the opposite direction. In this situation, it is true that I believed that my gate was 34B —my taking a particular directed action to move towards the higher-numbered gates illustrates that—but I believed it *just momentarily*.

[30] It's not even obvious that we must deny that one has the relevant belief over the extended interval. This scenario might best be understood as a case of conflicting belief instead. If one judges that *p* at *t*, one must also believe that *p* at *t*. But one may also judge that *p* at *t* while believing that ¬*p* as well.

(Schwitzgebel 2002, 2012; Cassam 2014, pp.117-119).[31] If we accept that belief is a complex structure of dispositions towards behavior, reasoning, and mental phenomenology, then judgment may not be sufficient for belief. While a judgment that *p* may sometimes manifest the dispositional structure that is belief, at other times a judgment might occur without the presence of the corresponding dispositional state.

Here we cannot do justice to the metaphysical debate about the nature of belief in its entirety. But I will briefly argue that we should accept that judgment is sufficient for belief even if we endorse a dispositional theory of belief.

The main condition of adequacy on any view of belief—dispositional or not—is that it capture the sense in which belief is a state of *taking to be true* (to put it roughly). To construct a dispositional notion of belief on the general level, we might consider whether we should ordinarily include a disposition to assert that *p* in the dispositional complex that is belief that *p*. How might we decide that? We should consider whether having a disposition to assert that *p* implies (at least defeasibly) that you take it to be true that *p*.

It is also the main condition of adequacy on an account of *judgment* that it capture the sense in which to judge that *p* is to take it to be true that *p*. We might give a dispositional account of judgment as well as a dispositional account of belief—an account on which an action's being a judgment is a matter of the dispositions it causes or manifests. If so, I cannot see any reason to label some set of dispositions that are together sufficient for *judgment* as *insufficient* for (at least momentary) belief. If both judgment and belief must capture a specific notion of *taking to be true*, then there should be an unbreakable implication from judgment to contemporaneous belief. This implication is entirely compatible with the possibility of judgment and lack of corresponding belief over time as well as the possibility of belief with little or no disposition towards judgment.

In principle we could also combine a dispositional analysis of belief with a different kind of individuation of judgment among mental actions. In this case, it would yet still be strange if judgment turned out to be insufficient for contemporaneous belief. The same fundamental characterization that judgment and belief share (their status as takings-to-be-true) would have to be respected in this kind of mixed analysis. On any view of the relationship between belief and judgment, then, judgment should come out to be sufficient for at least contemporaneous, momentary belief.

Recall that in judging that *p* in the course of the transparency method, you also *at the same time* self-attribute the belief for which that judgment is sufficient—namely, the belief that *p*. Thus, due to the sufficiency of judgment for contemporaneous belief, any transparent self-attribution of belief is guaranteed to be true.

3.2 Three aspects of warrant

3.2.1 Warrant for *self*-attribution

In using the transparency method to self-attribute beliefs, one kind of error is impossible: the error you would make by thinking that someone else's belief is your own. This type of error is impossible because use of the transparency method involves no identification

[31] I'd like to thank an anonymous reviewer for highlighting this point.

of yourself or anyone else. For this reason, the transparency method enjoys what Shoemaker (1968) has labeled "immunity to error through misidentification."[32]

It is not the case here that there is a particular ground that serves as justification for attributions of beliefs to yourself as opposed to others. Instead, the fact of the matter is that you cannot err in this way, and so you are entitled to the *self*-attribution in question given that you have the first-personal concept. Since judging *I believe that p* at all requires having that concept, any user of the transparency method is guaranteed to have entitlement to some such *self*-attribution of belief.[33]

3.2.2 Warrant for self-attribution of a *belief*

One crucial part of the warrant involved in transparent self-attributions of *belief* comes down to an agent's contrastive awareness of what she is doing in making judgments rather than, say, forming hopes or making suppositions for the sake of argument.

Contrastive awareness is the action awareness discussed in Section 2.1 above: the basic awareness the agent has of what she is doing when she is making judgments rather than forming hopes or making suppositions. It is guaranteed for any user of the transparency method, since using the transparency method involves performing intentional mental actions, and doing *that* involves having such contrastive awareness.

The other crucial aspect of warrant for transparent self-attribution of *belief* has to do with what is involved in having the concept BELIEF. It is trivially true that you have to have that concept to use the transparency method, since it involves thinking of things as beliefs. It is not trivial, however, that part of the understanding you must have if you have the concept BELIEF entitles you to apply that concept in the course of using the transparency method.

Part of what it is to have the concept BELIEF is to constrain yourself to use the mental action that is *judgment* (perhaps not thus conceptualized, by you) in using the transparency method. That implies that you would not use other mental actions in the same context, and that—given certain idealized conditions, other concepts, and capacity for more sophisticated reflection—you would reject as inappropriate the use of any other mental action in the place of judgment. It is this aspect of having the concept BELIEF that makes it the case that you are entitled to apply the concept in the transparency method. Since the transparency method requires the concept BELIEF, any user of that method is thus entitled to self-attribute a belief.[34]

To see why it is important to have this additional conceptual entitlement in warranting a self-attribution of a belief, we can consider what it would look like if someone lacking the concept BELIEF—and thus lacking the implicit understanding that entitles its application in the transparency method—tried to use the transparency method. Consider an agent called "Erraticus." Suppose that Erraticus has the same control over his mental actions, and the same contrastive awareness that that control implies, as the rest of

[32] See also Wittgenstein (1958), Shoemaker (1968), and Pryor (1999).

[33] For more on using the first-person in self-attribution of belief, see Boyle (2009b, pp.153-4).

[34] Here I do not mean to endorse the *general* principle that any inference or application of a concept whose availability to the subject is required for possessing that concept are inferences or applications to which the possessor is then entitled at any time. In its general form, this principle has interesting counterexamples. See Boghossian and Williamson (2003) for extended discussion.

us do. Now further suppose that Erraticus often attempts to use the transparency method for belief and fails, because he does not engage in judgment as the relevant embedded mental action. For example: sometimes, when asked for his belief whether *p*, he'll start making suppositions, come to a supposition that *p*, and say: "I believe that *p*."

Now, it should be clear that Erraticus could, in fact, stumble on the right way to use the transparency method as a matter of mere accident. In one instance, he might actually try to self-attribute what he calls a belief by way of making judgments. But even if Erraticus did this, he would not be *warranted* in what he took to be his self-attribution of 'belief,' because he clearly does not have the concept BELIEF. His lack of conceptual understanding is revealed by his erratic attempts and failures to use the transparency method for belief.

To recognize that it is Erraticus's lacking the concept BELIEF that matters to his lack of warrant in using a pseudo-transparency method is also to see that your having the concept BELIEF matters to your having warrant in using the actual transparency method for belief. Thus an account of the warrant you have for making a self-attribution of a *belief*, rather than any other attitude, must make essential reference to the entitlement that your conceptual understanding bestows on your use of the transparency method.

Together contrastive awareness and conceptual entitlement account for the warrant involved in transparent self-attribution of *belief*.

3.2.3 Warrant for self-attribution of a *particular* belief (that *p*)

A transparent self-attribution of a belief that *p* rather than, say, a belief that *q* is warranted due to the consciousness of the contents of intentional judgments. When your judgment that *p* is an intentional mental action, your consciousness has (at least in part) the content *p*. That consciousness of *p* in making that judgment is all that is needed by way of warranting a transparent self-attribution of a belief that *p* rather than any other belief.[35] This warrant is guaranteed for any agent using the transparency method as well: to use the transparency method is to perform an intentional mental action of judgment, and all such intentional mental actions have conscious contents.

This warrant should not be seen as evidential or inferential justification. The point is not that some relation of yours *to* your own consciousness that *p* (in judging that *p*) provides support for your self-attribution of the belief that *p*. Instead, your consciousness just *constitutes* your awareness that it is *p* that you believe.

3.3 Belief

[35] Those who take content externalism to threaten self-knowledge (e.g. Boghossian 1989) may disagree that this is all that's necessary by way of warrant here. I take the line endorsed by Burge (1996), Heil (1988), and Peacocke (1996) on this point: there is no such threat. Those still concerned about content externalism should at least note one nice feature of any given embedded mental actions: one and the same intentional mental action cannot enjoy two distinct environments that might contribute to the individuation of content. Cf. Burge (1996) on self-verifying judgments.

What remains to be explained is the formation of *belief* in a transparent self-attribution.[36] In using the transparency method, one *makes* a self-attribution of a belief that *p in judging that p*. To make a self-attribution of some belief is just to judge that one has that belief. And judgment is, as argued above, sufficient for contemporaneous belief. So in using the transparency method, one comes to believe that one believes that *p* just when one makes a judgment that *p*.

3.4 Sufficiency

The explanation up to this point demonstrates how transparent self-attributions meet three necessary conditions on knowledge: truth, warrant, and belief. But Gettier (1963) has shown that these are not generally jointly sufficient for knowledge.

It has proven notoriously difficult to fill the gap between warranted true belief and knowledge. Fortunately that is not necessary here. All that's necessary to show that the sort of knowledge at issue here cannot be Gettierized is to show that the warrant at issue cannot fail to be the right sort of warrant. But to use the transparency method at all is to have all of the right components of warrant for one's transparent self-attribution. That is because, as I have argued in Sections 3.2.1-3.2.3 above, all these components of warrant are guaranteed for a user of the transparency method. Hence this explanation is sufficient as an explanation of the knowledge status of transparent self-attributions.

4 Taking stock

In this paper thus far I have set forth objections to past interpretations of the transparency method and its epistemic structure. I have introduced the notion of embedded mental action, and then used that notion to provide an epistemic explanation of the knowledge status of transparent self-attributions of belief. How does this explanation stack up against those discussed earlier? What are its advantages, and what are its disadvantages?

4.1 Comparison with other explanations of the transparency method

The proposal of this paper owes a good deal to Moran's (2001, 2003, 2004) foundational insights. The deliberative authority you have with respect to your own beliefs remains fundamental to the epistemic explanation of the transparency method. Indeed, the key fact that embedded mental actions can belong to more than one type of mental action depends crucially on the fact that you have the agency to *intentionally* make up your mind on some matter (that is, make judgments).

However, the proposal presented here has two distinct advantages over Moran's epistemological analysis of the transparency method. This new proposal identifies *mental action awareness* as the basic self-awareness that contributes to warrant for transparent self-attributions of belief, and provides reason to think that we do indeed have such awareness. Moran's view, on the other hand, seems to require that you be aware of your own considerations for and against belief *as such* without providing much reason to think you have such awareness in the relevant cases of transparent self-attribution of belief.

[36] Some might argue that this is not even necessary; true warranted *judgment* could count as knowledge just as much as true warranted *belief* could. I simply accept this point.

Second, the view presented here fares better than Moran's with respect to explaining first-person authority. The transparency method is *infallible* in its production of known self-attributions of (contemporaneous) belief—but only in the first person.

The main goal of this paper, however, is to make better sense of the invaluable realization from Boyle (2009a, 2009b, 2011), Crane (2001) and Setiya (2011) that using the transparency method involves making no *transition* between *p* and *I believe that p*. Only by making sense of this suggestion is it possible to sidestep the problems about inferential justification faced by the Gallois-Byrne proposal. The notion of embedded mental action is meant to do just that.

Setting aside the notion of embedded mental action, the proposal set out in this paper bears certain undeniable affinities to other no-move views. Crucially, Boyle (2011) has highlighted the need to understand the role of action awareness in self-knowledge. And Setiya (2011) has noted in passing that, in using the transparency method, "Evans's procedure is performed *in anticipation*" (p.187, emphasis added). The discussion of this paper can help spell out what it means to do that: it involves antecedent conceptualization of the contents of your upcoming judgments as the contents of your beliefs.

4.2 Advantages

The proposal to understand the transparency method for belief as involving embedded mental action has two important advantages: first, it can explain epistemic and psychological immediacy; and second, it can be generalized in a principled way to explain when and how transparent self-knowledge of *other* attitudes is possible.

4.2.1 Immediacy

The intuition that transparent self-knowledge of belief is both psychologically and epistemically immediate is a natural and influential one.[37] On the view presented here, transparent self-attributions of belief are immediate in both ways.

Self-knowledge of belief, when produced by the transparency method in the way discussed above, is *psychologically* immediate in that there is absolutely no space—that is to say no reasoning, and no inference—between judging that *p* and self-attributing the belief that *p*. On the contrary, the two mental actions are one and the same in this context.

Self-knowledge of belief in this context is *epistemically* immediate in that your warrant for believing that you believe that *p* is not itself dependent upon your warrant for any of your other beliefs (cf. Pryor 2003). Your warrant for any given belief that you self-attribute via the transparency method might be severely lacking. The poverty of *that* justification, though, would not weaken your warrant for *self-attributing* the belief in question when you do so by using the transparency method. That is just as it should be: a poorly warranted or unwarranted belief is no less a belief than a fully warranted one.

4.2.2 Potential for extension

[37] See Cassam (2014), p.5-6, for a helpful characterization of these intuitions. Compare Moran (2001), Section 4.5, and Heal (2002), p.2. Russell's (1912) claim that we are directly acquainted with our own mental states can also be understood as expressing the thought that our knowledge of our own beliefs is both epistemically and psychologically immediate.

The explanation of the transparency method presented above can be generalized to explain when and how other attitudes or mental actions can be transparently self-known.

Three central facts about belief stood out as crucial to the epistemic explanation of the transparency method. In generalized form, these facts provide a set of sufficient conditions which, when met, ensure that a mental attitude or action *M* can be transparently self-known in a way analogous to belief.

If some mental state or action *M* (with propositional content *p*) meets the following three conditions, then it will also be transparently knowable to its subject:

1. There is a kind of occurrent mental action *a* with content *p* the intentional performance of which is sufficient for the contemporaneous presence of *M* with content *p*;
2. *a*'s being an action of that kind (the kind sufficient for some such *M*) is tracked by at least basic contrastive mental action awareness guaranteed in the intentional performance of *a*; and
3. Having the *M*-concept entitles an agent to its application in the performance of an *a* in a transparency method for *M*

What is a transparency method for any such *M*? To use some such transparency method for some *M* is just to do all of the following. First, you intentionally set out to self-attribute an *M* meeting some content requirement. As a result, you intentionally set out to perform an action of type *a* (perhaps not under a particularly conceptually rich description) with some content that meets the relevant requirement, already understanding the content of that *a* as the content of some *M*. As a result of *that*, you actually do go on to perform some such *a* with a content that meets the requirement on the *M* to be self-attributed—and *in so doing*, you judge of yourself that you have an *M* with that content. Call any such self-attribution a *transparent self-attribution of an M*.

For reasons precisely parallel to the ones discussed in Section 4 above, any *M* meeting the three relevant conditions can be self-attributed in this way—and the resultant transparent self-attribution will have the status of knowledge. It will be true because performing some *a* with content *p* is sufficient for having (or performing) an *M* with content *p*; it will be warranted due to immunity to error through misidentification, contrastive action awareness, and conceptual entitlement; and it will be judged, or believed, because the *a* in question will be an embedded mental action in the context of the relevant transparency method. Any such *a* will also be a self-attribution of an *M*. The warrant involved in the relevant transparency cannot be Gettierized.

Is there any such *M* other than belief that meets the three conditions in question? Intention is a promising candidate. You can decide to do something as an intentional mental action (as in practical deliberation), and actually deciding to do something is sufficient for having the intention to do it.[38] In being aware that you are doing *this sort of thing* (deciding what to do), you can also be aware that you have a particular intention.

4.3 Objections and replies

[38] See, however, Paul (2012) for an argument to the contrary.

4.3.1 The rarity of active self-reflection

A critic might question how often we really embark on this project of intentionally directed thought in order to probe our own beliefs. If the answer is "not often," it might be thought that this account cannot do all that much to explain the full scope of privileged access, or of first-person authority. You are authoritative, the critic might claim, over far more of your own beliefs than those that you transparently self-attribute.[39]

It seems to me that we do, quite often, use the transparency method—at least as often as we explicitly consider the question of what we ourselves believe. Though its philosophical explanation is somewhat complicated, actually using this method can be so simple as to be practically mindless. It doesn't require deliberation of any sort—a judgment that *p* can express, rather than form, a belief—and you need not formulate anything complex to yourself in order to do it.

Yet the transparency method need not *actually* be used all that often in order to explain the full scope of privileged access or first-person authority. To explain your privileged access to some state such as belief, we need only to explain why you have a method you *could* use, at any point, to self-attribute a belief such that the self-attribution in question is more likely than any third-personal one to amount to knowledge. A similar point applies for first-person authority: to say you are an authority on what you believe is not necessarily to say that you often explicitly consider your beliefs as such. All that the claim of first-person authority implies is that, *were* you to consider what you believe, your word would trump anyone else's word on the matter. For these reasons, the understanding of the transparency method presented in this paper still offers a powerful way to make sense of both privileged access and first-person authority.

4.3.2 Diachronic belief and fallibility

The transparency method produces infallible self-attributions for the synchronic case, but—since judgment at some moment *t* is *not* sufficient for belief over any extended interval of time—it cannot produce infallible self-attributions for the diachronic case. A critic might worry that the transparency method may therefore be completely unable to explain any self-knowledge of *diachronic* beliefs. And very often it is diachronic belief that we care about, rather than what you believe at just one particular moment in time.[40]

It is undeniable that the transparency method as explained above cannot ensure perfect knowledge of diachronic beliefs, and it is important to note the potential for error in the transparency method as applied to these diachronic cases. Still, neither of these concessions implies that the transparency method can do nothing to explain self-knowledge of diachronic belief. Why couldn't you use the transparency method as a *fallible* method for attribution of diachronic belief as well?

Of course there will be differences in the epistemology of the transparency method for diachronic belief. But given an independent theory of the warrant required for

[39] I'm indebted to an anonymous reviewer of this paper for a careful presentation of this objection.

[40] I'm grateful to [ACKNOWLEDGEMENT REDACTED FOR ANONYMITY] for expressing this point in a compelling way.

knowledge, surely *some* uses of the transparency method can contribute to production of diachronic self-knowledge of belief. Perhaps more warrant will be needed in such cases than that described in Section 3.2 above. Still, there is no reason in principle why the understanding of the transparency method advanced in this paper cannot be extended to make significant headway towards understanding self-knowledge of diachronic belief.

5 Conclusion

Understanding the transparency method in terms of embedded mental action seems to be the best option available. This interpretation of the method makes available a full epistemic explanation of the transparency method. This explanation makes sense of the psychological and epistemic immediacy and first-person authority of transparent self-knowledge of belief. The explanation provided here can also be extended in a principled way to make sense of other transparent self-knowledge we might have—including knowledge of intention, and knowledge of diachronic attitudes.

References

Anscombe, G.E.M. (1957). *Intention*. Oxford: Basil Blackwell.

Boghossian, Paul (1989). "Content and self-knowledge." *Philosophical Topics* 17.1: 5-26.

Boghossian, Paul (2014). "What is inference?" *Philosophical Studies* 169: 1-18.

Boghossian, Paul and Timothy Williamson (2003). "Blind reasoning." *Proceedings of the Aristotelian Society* 77: 225-293.

Boyle, Matthew (2009a). "Active belief." *Canadian Journal of Philosophy* 39 sup. 1: 119-147.

Boyle, Matthew (2009b). "Two kinds of self-knowledge." *Philosophy and Phenomenological Research* 78, 133–63.

Boyle, Matthew (2011). "Self-knowledge and transparency II: Transparent self-knowledge." *Proceedings of the Aristotelian Society* Supp. Vol. LXXXV: 223-241.

Brueckner, Anthony (1998). "Moore inferences." *The Philosophical Quarterly* 48.192: 366-369.

Burge, Tyler (1996). "Our entitlement to self-knowledge I." *Proceedings of the Aristotelian Society* 96: 91-116.

Byrne, Alex (2005). "Introspection." *Philosophical Topics* 33.1: 79-104.

Byrne, Alex (2011). "Self-knowledge and transparency I: Transparency, belief, intention." *Proceedings of the Aristotelian Society* Supp. Vol. LXXXV: 201-221.

Byrne, Alex (2012). "Knowing what I want." In Jeeloo Liu and John Perry, eds., *Consciousness and the Self: New Essays* (pp. 165-183). New York: Cambridge University Press, 2012.

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

Crane, Tim (2001). *Elements of Mind: An Introduction to the Philosophy of Mind*. New York: Oxford University Press.

Davidson, Donald (1963/2001). "Actions, reasons, and causes." In Donald Davidson, *Essays on Action and Events* (pp.3-19). Oxford: Clarendon Press.

Davidson, Donald (1973/2001). "Freedom to act." In Donald Davidson, *Essays on Action and Events* (pp.63-81). Oxford: Clarendon Press.

Davidson, Donald (1984/2001). "First person authority." In Donald Davidson, *Subjective, Intersubjective, Objective* (pp. 3-14). Oxford: Clarendon Press.

Davidson, Donald (1987/2001). "Knowing one's own mind." In Donald Davidson, *Subjective, Intersubjective, Objective* (pp. 15-38). Oxford: Clarendon Press.

Edgley, Roy (1969). *Reason in Theory and Practice*. London: Hutchinson.

Evans, Gareth (1982). *The Varieties of Reference*. New York: Oxford University Press.

Frege, Gottlob (1979). "Logic." In Gottlob Frege, *Posthumous Writings* (pp.1-8). Oxford: Basil Blackwell.

Gallois, André (1996). *The World Without, the Mind Within*. New York: Cambridge University Press.

Gettier, Edmund (1963). "Is justified true belief knowledge?" *Analysis* 23.6: 121-123.

Green, Mitchell and John N. Williams, eds. (2007). *Moore's Paradox: New Essays on Belief, Rationality, and the First Person.* New York: Oxford University Press.
Heal, Jane (2002). "On first-person authority." *Proceedings of the Aristotelian Society* 102: 1-19.
Heil, John (1998). "Privileged access." *Mind* 97: 238–251.
Moran, Richard (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
Moran, Richard (2003). "Responses to O'Brien and Shoemaker." *European Journal of Philosophy* 11.3: 402-419.
Moran, Richard (2004). "Replies to Heal, Reginster, Wilson, and Lear." *Philosophy and Phenomenological Research* 69.2: 455-472.
O'Brien, Lucy (2003). "Moran on agency and self-knowledge." *European Journal of Philosophy* 11.3: 375-390.
O'Brien, Lucy (2005). "Self-knowledge, agency, and force." *Philosophy and Phenomenological Research* 71.3: 580-601.
O'Brien, Lucy (2007). *Self-Knowing Agents*. New York: Oxford University Press.
O'Brien, Lucy and Matthew Soteriou, eds. (2009). *Mental Action*. New York: Oxford University Press.
Paul, Sarah K. (2012). "How we know what we intend." *Philosophical Studies* 161: 327-346.
Peacocke, Christopher (1996). "Our entitlement to self-knowledge II: Entitlement, self-knowledge, and conceptual redeployment." *Proceedings of the Aristotelian Society* 96: 117-158.
Peacocke, Christopher (1998). "Conscious attitudes, attention, and self-knowledge." In Crispin Wright, Barry C. Smith, and Cynthia MacDonald, eds., *Knowing Our Own Minds* (pp.63-121). Oxford: Clarendon Press.
Peacocke, Christopher (2008). "Mental action." In Christopher Peacocke, *Truly Understood* (pp.245-285). New York: Oxford University Press.
Pryor, James (1999). "Immunity to error through misidentification." *Philosophical Topics* 26.1: 271-304.
Pryor, James (2005). "There is immediate justification." In Matthias Steup and Ernest Sosa, eds., *Contemporary Debates in Epistemology*. Malden, MA: Blackwell.
Russell, Bertrand (1912/2013). *The Problems of Philosophy*. Online: Global Grey.
Ryle, Gilbert (1949). *The Concept of Mind*. London: Barnes & Noble.
Ryle, Gilbert (1971a). "A puzzling element in the notion of thinking." In Gilbert Ryle, *Collected Papers, Volume II: Collected Essays 1929-1968* (pp.391-406). London: Hutchinson.
Ryle, Gilbert (1971b). "Thinking and reflecting." In Gilbert Ryle, *Collected Papers, Volume II: Collected Essays 1929-1968* (pp.465-479). London: Hutchinson.
Ryle, Gilbert (1971c). "The thinking of thoughts: What is 'Le Penseur' doing?" In Gilbert Ryle, *Collected Papers, Volume II: Collected Essays 1929-1968* (pp.480-496). London: Hutchinson.
Schwitzgebel, Eric (2002). "A phenomenal, dispositional account of belief." *Noûs* 36.2: 249-275.
Schwitzgebel, Eric (2012). "Self-ignorance." In JeeLoo Liu and John Perry, eds., *Consciousness and the Self: New Essays*. New York: Cambridge University Press.
Setiya, Kieran (2011). "Knowledge of intention." In Anton Ford, Jennifer Hornsby, and Frederick Stoutland, eds., *Essays on Anscombe's Intention* (pp.170-197). Cambridge, MA: Harvard University Press.
Shah, Nishi and J. David Velleman (2005). "Doxastic deliberation." *The Philosophical Review* 114.4: 497-534.
Shoemaker, Sydney (1968). "Self-reference and self-awareness." *Journal of Philosophy* 65.19: 555-567.
Shoemaker, Sydney (1988). "On knowing one's own mind." *Philosophical Perspectives* 2: 183-209.
Shoemaker, Sydney (2003). "Moran on self-knowledge." *European Journal of Philosophy* 11.3: 391-401.
Silins, Nicholas (2012). "Judgment as a guide to belief." In Declan Smithies and Daniel Stoljar, eds., *Introspection and Consciousness* (pp.295-327). New York: Oxford University Press.
Williams, Bernard (1976). "Deciding to believe." In Bernard Williams, *Problems of the Self: Philosophical Papers 1956-1972* (pp.136-151). Cambridge: Cambridge University Press.
Wittgenstein, Ludwig (1958). *The Blue and Brown Books*. Oxford: Blackwell.
Wright, Crispin (1989). "Wittgenstein's later philosophy of mind: Sensation, privacy, and intention." *Journal of Philosophy* 86.11: 622-634.