Abstract: Self-consciously attempting to shape one’s beliefs through deliberation and reasoning requires that one stand in a relation to those beliefs that might be signaled by saying that one must inhabit one’s beliefs as one’s own view. What does this amount to? A broad swath of philosophical thinking about self-knowledge, norms of belief, self-consciousness, and related areas assumes that this relation requires one to endorse, or be rationally committed to endorsing, one’s beliefs. In fact, however, fully self-conscious adherence to epistemic norms requires the ability to self-consciously hold a belief without endorsing that belief as true, as well-supported by the evidence, or as meeting some other epistemic standard, and there are cases in which no such commitment is rationally required. This ability is necessary if there is to be any such thing as a fully self-conscious process of changing one’s mind.

Keywords: belief, self-consciousness, transparency, epistemic norms, self-knowledge, first-person

1. Introduction
Mature adults have the capacity to self-consciously follow epistemic norms in inquiry and deliberation. As I will argue, this capacity requires the possibility of self-consciously holding a belief without endorsing that belief as true, as well-supported by the evidence, or as meeting any other epistemic standard. Otherwise, there couldn’t be any such thing as a fully self-conscious process of changing one’s mind.

While this claim is of considerable interest in its own right, it also raises important doubts about several significant philosophical agendas.

A broad tradition of thought urges that what makes an attitude properly one’s own in some sense deeper than mere correctness of ascription from an external perspective – one’s own as subject, we might say – is one’s endorsement of it. On the line of thought I will urge, however, endorsement does not forge a distinction between what properly belongs to “me” as subject and that in my psychology to which I relate only as an object of description, report, and possibly management.

Several recent strands of thought have urged that rational endorsement, rational commitment, and the related ability to self-ascribe an attitude through consideration of the matters the attitude is about are crucially important for understanding distinctively first-personal self-knowledge of our beliefs and other attitudes (Bilgrami 2006, Boyle 2009, Evans 1982 p.225, Moran 2001, Rödl 2007). The considerations that I will put forward likewise suggest that such projects start off on the wrong foot. As I will argue, fully self-conscious belief does not constitutively require or commit one to the ability to self-ascribe in accordance with Evans’s Transparency Procedure. Nor does it require or commit one to any positive epistemic evaluation or endorsement of one’s attitude.
There is a deeper issue here. Kant held that “every conscious combination presupposes unity of consciousness, and consequently a rule for the combination” (Anthropology 7:137-138). I will argue, however, that the combination of attitudes in a single self-conscious perspective need not be in accordance with rules. The unity of a self-conscious perspective need not be a rational unity. In fact, I will argue that it is our capacity to self-consciously occupy rationally incoherent states of mind that makes self-conscious, active deliberative rationality possible.

I begin (section 2) by identifying the relation to one’s beliefs that is required for self-conscious inquiry and deliberation. A broad and influential approach takes this relation to constitutively involve or normatively require some form of endorsement (section 3). Familiar considerations about our relation to our attitudes more generally, particularly our emotions, raise questions about this approach (section 4). In fact, familiar epistemic phenomena tell quite directly against it (sections 5 and 6). These phenomena are closely related to self-conscious change of mind, and self-conscious revision of one’s views is impossible without the capacity to believe self-consciously while withholding endorsement of one’s belief (section 7). These results are not simply negative. They point the way to a different framework and set of starting points for philosophical theorizing in these areas (section 8).

2. The Phenomenon at Issue

It is widely agreed that there is such a thing as reflectively making up one’s mind and thereby shaping one’s beliefs – a process of determining what one believes by determining what to believe – and that such processes of active deliberation can take
place in a fully self-conscious way.\(^2\) Self-conscious deliberation requires that one stand in a distinctive stance or relation to one’s beliefs, one often termed “first-personal” as distinct from “third-personal” or “merely attributional”. An essential first task is to characterize this stance or relation in a way that does not beg any crucial questions.

Suppose that you self-consciously engage in a process of reasoning. You begin with premises that you recognize as part of your current view, and you understand yourself to be drawing a conclusion from them – adding something to your view – in accordance with principles which you accept. Doing this requires more than mere knowledge about your beliefs. Someone who learns about his repressed belief from a trusted therapist or self-observation does not yet stand in the requisite relation to it, though he knows about it. Even if he learns about some psychological process going on in him involving transitions from this belief to other mental states, he would not understand himself to be thinking thereby about how to enrich his current view by moving from premises he accepts to further conclusions. Whatever else it might be, this would not be an episode of self-conscious deliberation.

The shortcoming here is not aptly identified by the standard characterizations of what is epistemically distinctive about first-personal psychological self-knowledge. We can imagine someone equipped with a cognitive mechanism that yields immediate, non-observational and non-inferential knowledge about his repressed beliefs. We can add that this mechanism connects with informational capacities enabling direct first-personal self-reference in a way that precludes error about who has these beliefs.\(^3\) We can give the person all of the relevant conceptual capacities. Still, he might not be enabled thereby to understand himself to be engaged in deliberation from these beliefs as premises. Despite
these additions, his relation to these beliefs appears to be no different in the crucial respects than in the case of simple repression. The epistemic terms thus don’t fully capture the distinctive way in which one relates to one’s belief in self-conscious reasoning and deliberation; even if (as I doubt) someone who meets the epistemic characterizations must thereby relate to her beliefs in that distinctive way, this characterization doesn’t tell us what that way is.4

The person we are imagining can self-ascribe beliefs using the first-person pronoun, but he cannot unite those self-ascriptions and the beliefs they are about into a single viewpoint because the beliefs are repressed: the person cannot occupy their subjective standpoint as conscious subject, consciously “seeing” the world in the way they involve. When one has a conscious belief, by contrast, one can occupy the subjective perspective of the belief as a conscious subject. One then sees the world through that lens, as it were, and the content of the belief characterizes something of how one consciously takes things to be. Taken together, one’s conscious beliefs form a single evolving view that includes a second-order view on one’s perspective on the world, including self-ascriptions of beliefs by means of the first-person pronoun. In some cases – as when one knows about one’s repressed beliefs – these self-ascriptive beliefs concern first-order beliefs which are not part of one’s total conscious view. In other cases, the ones centrally relevant here, they concern beliefs which are part of one’s total conscious view, and so both the first-order and second-order aspects of one’s view are part of a single, complex subjective perspective. In these cases, the self-ascription is made from within the same total conscious subjective perspective that is partially constituted by the beliefs it is about.
Self-conscious deliberation and inquiry in accordance with norms require that one relate to one’s beliefs in precisely this way. The person must be able to integrate her thinking in a way that enables her to think about her view as her own view even while occupying and operating from within it. To do that, she must be able to self-ascribe the attitudes from within the very same standpoint whose perspective they partially constitute. I will use the phrase “self-conscious belief” to refer to beliefs to which one relates in this way.

Self-conscious belief, characterized in this way, is not only essential for self-consciously inferring a conclusion from premises, but also for self-conscious attempts to modify, correct, or update one’s current beliefs. How else could I understand myself as engaged in a process of revising my own view? Consider, too, what happens when we recognize a contradiction amongst our beliefs. Why do I take this conflict to matter to me in such a way as to call for my deliberative response? Only because I understand this all to be part of my current view from the subjective vantage point of the holder of that very view. Ordinary inquiry likewise sometimes requires recognition that one lacks crucial information. But one cannot recognize anything as an informational gap in one’s own view unless one is capable of self-consciously occupying the subjective position of that view.

In short: In order to apply epistemic norms in the course of self-conscious deliberation and inquiry, one must be able to apply those norms from the vantage point of the very person whose view it is – from the subjective position of that very view – in the shaping of that view. For that to be possible, one must be able to occupy the subjective
position of one’s view, understanding it precisely as one’s view. This is the phenomenon that we need to understand.

3. One Approach: Variations on a Theme

What is involved in relating to one’s beliefs in this way? A wide variety of philosophers take a stance of endorsement to be central here. Consider, for instance, the underlying conception shared by the following passages.

"Consciously believing that P commits one, on reflection, to … epistemically endorsing one’s own belief that P” (Huemer 2008, 148).

“From the first-person point of view, I must take what I believe about something to be the expression of my sense of the reasons relating to the content of that belief” (Moran 2012, 232).

Both authors here take self-conscious belief to involve some form of positive stance towards one’s belief. Beatrice Longuenesse explains Kant’s “I think” in a way that makes a similar presumption.

"suppose I say, ‘This is a tree,’ and someone asks: ‘Are you sure?’ I might reply: ‘Yes, I think this is a tree.’ ‘I think’, here, does not express a turning back of my attention on to me. It’s just the expression in words of the fact that I can stand by my statement because I’ve been through the drill of combining the pieces of sensory information available to me, comparing them to others, forming recognitional concepts and eventually coming up with the proposition, ‘This is a tree,’ which is now available as a premise for further reasoning.” (2012, 89-90, italics added)

Wittgenstein sounds a related theme. “One can mistrust one’s own senses, but not one’s own belief” (Philosophical Investigations, pt. II, p. 190), he writes, urging that no sense can be made of a present-tense self-ascription of false belief (PI, pt. II, p. 162). And Jane Heal uses such considerations to motivate a challenge to functionalist views of belief:
if functionalism does not deliver a contradiction [in the self-ascription, “I believe that p, but not p’”] we can have the subject acknowledging the supposed belief but also disowning it. And then the functionalist story has not captured the ordinary sense of ‘believe’ (1994, 19).

As her use of the word “disown” indicates, Heal here presumes that one cannot simultaneously self-ascribe a belief from the subject position, acknowledging it as one’s own, and also regard it as false.

These are just a few examples of a broad current of thought which takes some form of endorsement to be required here, so that self-attribute of a belief from the subject position must involve a stance of approval tied to considerations of truth, rationality, or positive epistemic status. This idea can be developed in various ways – for instance, as a normative rather than constitutive requirement, or as a requirement for actual endorsement rather than commitment to a positive evaluative stance – but it is the broad, underlying tendency, pitched at the level of maximal generality, that I want to question. For this purpose, it would be wrong to begin by sharply defining “endorsement”. Until the underlying tendency is addressed and an alternative is shown to be intelligible, no argument against particular theses will carry conviction. The underlying tendency will just reappear in a different form.

This broad conception manifests itself in various ways within the context of particular philosophical projects. In discussions of epistemic norms, for instance, it is sometimes held that self-conscious belief constitutively requires one to regard -- or, alternatively, rationally commits one to regarding – one’s belief as meeting some epistemic standard such as being a product of the proper operation of one’s rational faculties, supported by the evidence, warranted or justified, or even knowledge. Such a view is often supported by the consideration that it can seem irrational or incoherent if a
person avows a belief while acknowledging that it fails to meet some epistemic standard. Huemer, for instance, argues that knowledge is a constitutive norm of belief on the grounds that “I cannot rationally continue to believe … while reflectively refusing to epistemically endorse that belief” (2008, 148).

Another line of thought takes the very concept of belief to require that in self-attributing a belief from the subject position one thereby commits oneself to the truth of one’s belief (not just the truth of the believed proposition). Moran writes of “the commitment that beliefs I call my own are beliefs I can endorse as true,” claiming that “that commitment is internal to the very concept of belief” (2001, 105). Closely related here is the idea that to regard something as my belief is to regard a question as resolved (Hieronymi, 2006). Moran expresses this further idea when he writes, “for the person herself, if her belief that it is raining does not constitute the question’s being settled for her, then nothing does” (2001, 77).

The Transparency Condition likewise supposes that self-conscious belief constitutively involves an endorsing stance towards one’s belief. The fundamental idea here is that to answer a question about what one believes, one must attend to those relevant first-order considerations about the world that (readily) present themselves from within one’s current perspective: one can (under conditions of rationality) answer the question whether one believes that \( p \) by reflecting on the question whether \( p \) and then appending “I believe…” to whatever result one reaches. On this view, integration between one’s first-order and higher-order perspectives is brought about through one’s ability to avow as one’s belief the results of one’s first-order directed thinking. This suggests several possible necessary conditions on a self-conscious relation to one’s
beliefs. Perhaps the weakest would simply hold that it is a constitutive condition on self-attributing a belief that $p$ from the subject position that one must take oneself – or is thereby rationally committed to taking oneself – to be able to self-ascribe the belief by asking oneself whether $p$. But even this condition involves an endorsing stance. If in self-conscious belief about $X$ one must take oneself to be able to self-attribute the belief by considering the facts about $X$, then one must also take one’s belief to be true or supported by the relevant reasons/evidence.\(^8\)

4. An Initial Comparison

Consider attitudes other than belief. It is uncontroversial that a mature adult can feel an emotion such as anger, recognizing it as such, without taking any sort of approving stance towards it. I might say to my spouse, genuinely feeling anger as I do, “I am very angry with you about the way you’ve handled this. But I want to hear your side. How does this look to you?” I speak from a position of anger, but I also acknowledge that there is another perspective, and for the time being I hold in abeyance endorsement of my anger as appropriate or rational. Even if I find that I am wrong to feel angry, I may nonetheless continue to do so. Then I still speak from a position of anger, only now as someone who also recognizes that he is unreasonably angry. “I am sorry; I’m still angry. Give me time to cool off.” I will then attempt to prevent my anger from playing a role in my practical reasoning and action.

In this sort of case it would be wrong to say that one merely has an inclination to be angry, is inclined to be angry, or finds a feeling of anger within oneself. One might like to say such things, perhaps defensively, but they wouldn’t be true. One is angry.
One’s orientation may also include desires to be fair and to get things right, concern for the relationship and for the other person, recognition that one’s anger is out of place, and so on. That does not make it less true that one is angry. One’s overall orientation is complicated. To pretend otherwise is to lie to oneself.

Someone who can’t take up this sort of stance regarding his emotions and related attitudes is very difficult to interact with. He can’t simultaneously occupy his perspective and regard it as his perspective, one among many, genuinely holding open the possibility that it is not correct. We call such a person unreasonable. A pervasive incapacity of this sort doesn’t just make someone difficult to interact with; it’s a mark of severe personality disorder. To be incapable of occupying your perspective, recognized as your perspective, while holding open the real, live possibility that you are wrong is to have failed to achieve a crucial developmental milestone.⁹

There is no obvious reason to think that self-conscious belief is any different. Consider some of the key functional, dispositional, and phenomenal considerations to which an adequate theory of belief should, prima facie, be responsive.

(a) Third-person belief attribution is constitutively tied to ceteris paribus patterns and dispositions relating to judgment, reasoning, feeling, motivation, action, and assertion.¹⁰

(b) Occurrent self-conscious belief that p involves feelings of conviction, and someone whose belief is not repressed, for instance, is disposed to have such feelings (all else equal).

(c) Self-conscious occurrent belief also constitutively involves that p’s being the case is part of the subjective perspective through which and
against the background of which the person encounters the world. The world will be presented to and in her occurrent thought as including that \( p \) is the case, and this presentation will include a distinctive phenomenology, which we might term a sense or feeling of reality.\(^{11}\) Her experience will likewise be shaped in various ways by the ‘fact’ of \( p \), including for instance a tendency to interpret her experience in certain ways. We might say that in her subjective world, \( p \) is the case.

(d) Self-conscious belief that \( p \) includes various other phenomenal dispositions, too, such as a disposition to feel surprise upon suddenly learning otherwise and to have emotional responses congruent with \( p \).

(e) Self-conscious belief also involves various dispositions concerning reasoning and action. The person will be disposed to judge that \( p \) when she considers relevant questions. \( p \) will be presented in her conscious directed thinking as a premise for her reasoning about both what to do and what is the case, and she will be disposed to so deploy it. She will also be disposed to assert \( p \) when appropriate and to act on the basis of \( p \).

(f) These dispositions are \textit{ceteris paribus}, and many of them – especially the latter three -- are also importantly defeasible in a very particular way: they can be resisted by the person.

So far as these considerations go, a person could have a self-conscious belief that \( p \) and yet also not endorse that belief as true (say, because she regards that belief as false). For such a thing to happen, all that would be required is that the relevant patterns, dispositions, and phenomenal elements take a very particular -- and complex -- shape.
And it makes no difference if we add that belief involves what is, in some sense, a commitment to the truth of the relevant proposition, since one can recognize one’s commitments without endorsing them. One can find that one has a commitment that is entirely misplaced.

What of the contention that in addition to the above characteristics, a stance of endorsement towards one’s attitude is integral to the notion of belief itself? Here I suggest an open mind. Let’s see what fits best with the phenomena overall.

5. The Argument from Double-Checking and Discovery of Error

Two everyday epistemic phenomena are particularly illuminating here.

A. Discovery of error

We sometimes discover with full self-conscious awareness that our beliefs are false. We are quite sure that things are a certain way, and then even while thinking that things are that way, we are faced with the fact that this isn’t so. This phenomenon is perhaps most prominent when one has significant investment – whether emotionally or because of the weight of conviction – in being right. Think, for instance, about the experience of finding that a beloved friend has fundamentally betrayed you. The belief that this person would never do you harm can be present in your conscious thinking even as you recognize its falsehood. In such cases, one can occupy the subjective position of one’s belief, viewing the world in the way it involves, even while recognizing that one’s view is false. The particular interpersonal aspects of this example are not essential to the phenomenon. My spouse, for instance, recounted an experience in which she faced the realization that certain people had intentionally acted to destroy another person’s career.
This discovery clashed with a view of humanity that she held dear. She realized in that moment that her view was incorrect. She then had to adjust to her discovery. Though not every case of belief change involves such a moment, this phenomenon can sometimes show up in less dramatic fashion when one recognizes that one’s view, understood as such, is decisively refuted by the evidence.

The significance of such moments is a commonplace outside of philosophy. The popular-science blogger Ethan Siegel writes,

“No matter who you are, no matter how smart you are... there will come an instance where the evidence you encounter will be irreconcilable with the picture of reality you presently hold. And when that moment happens... You have to open your self up to at least the possibility that you are wrong” (2012).

Similarly, on the comments page to Kathryn Schulz’s TEDtalk “On being wrong,” one finds this: “Really great talk, since I first saw it I've started to notice how I react to realising I am wrong and shaping the way I respond by accepting that I am wrong…”

As both comments highlight, such moments can be not only important, but also difficult in familiar ways.

One source of difficulty here may be a sting to one’s pride or sense of competence or authority. Another is the rationalizing force of belief perseverance, and a third the pain of recognizing that something unwelcome is the case. Underlying all of these, however, is the uneasiness that can arise as one comes to see the force of an alternative view. This is an experience of a moment of conflict between one’s view and reality - a
form cognitive dissonance – which is itself unpleasant and so something we are tempted to avoid.14

B. Double-checking and critical self-scrutiny

Even as I self-consciously believe that p, I might find some ground for suspicion that all is not well with my belief – that it is false, not supported by the evidence, shaped by irrational processes of bias and wishful thinking, or the like – but I might not yet have enough reason to revise my belief. It may then be appropriate for me to double-check whether my belief is true, supported by the evidence, or whatever – even as I self-consciously believe that p. (The existence of such cases requires just that one can fully appropriately believe that p even if one recognizes some weak evidence that points another way or suggests that all is not as it should be with one’s belief.15)

Here is an example. At Hanukah dinner, the host placed six candles in the menorah. A guest who had been keeping track of the candle lightings was certain that it was the seventh night. She asked, “Shouldn’t there be seven candles?” She then began counting off nights on her fingers to check whether she was right. As she reported afterwards, she did not cease to believe that it was the seventh night, nor did she somehow hedge or qualify her belief: she remained certain that it was the seventh night. But nor was she merely counting off nights in order to convince the host that it was the seventh night. Rather, she genuinely wondered whether she was right to be so certain; she was double-checking to see whether her belief was correct, seeking to vindicate her belief to herself. Given the situation, she did not take herself to have reason to revise her belief or suspend judgment. But she did have reason to double-check whether her belief
was true. In this case, her response was perfectly reasonable: even while believing that it was the seventh night, it was appropriate for her to check whether she was right.\footnote{Such cases are very common. Having carefully looked for your keys in your bedroom, you might be self-consciously certain that they aren’t there, and yet, not finding them elsewhere, you might reasonably double-check. You might genuinely wonder whether your belief is correct. Likewise, a detective might reasonably say, “I’m fully convinced that he’s guilty, but to be sure we are right we still need to check on that rumor that the witness was bribed.” Such phenomena are also familiar in scholarly life. One might believe one’s view on the basis of what one takes to be decisive reasons, and yet – recognizing that philosophy is hard, that one sometimes goes astray, and that one may be under the sway of a pet theory – one might seek out and consider possible counter-arguments in order to check whether one is right. And given a piece of apparent counter-evidence, one might reasonably continue to hold one’s belief but say, “I need to doublecheck my reasoning to see whether I’m right.” During the inquiry one avows the belief, self-ascribing it from the subject position, even while investigating – and genuinely wondering – whether one’s belief is true. This is not always a matter of failing to hold a certain higher-order belief. Someone may believe that her belief is true, that she knows that p, or that her belief that p is epistemically well-supported. These latter beliefs may themselves be the targets of reasonable double-checking.

There is no irrationality in any of these procedures. A person who has reason to double-check does not have reason to give up her belief, but she has sufficient reason to inquire into its truth or epistemic propriety – genuinely wondering whether her belief is
correct – without relying upon her antecedent commitment to the truth of p in the course of that inquiry. 17

C. Why Take These Phenomena At Face Value?

Prima facie, reasonable double-checking and discovery of error fit quite naturally with the points highlighted earlier about attitudes such as anger. If I can recognize that I am feeling angry even though anger isn’t called for, why can’t I equally recognize that I consciously believe that p even though belief that p is out of place? If I can genuinely wonder whether my anger is appropriate, why can’t I equally genuinely hold open the question of whether my belief is appropriate? A significant burden lies on the philosopher who wishes to understand these matters in a different way.

It might be suggested that when one discovers one is wrong, one is only inclined to believe, or it only appears to one, that things are otherwise. However, in that moment one’s relation to the thought that things are otherwise can be closer than that: it can be, for the moment, an aspect of one’s current view. This shows up in the first-personal aspects of the experience, particularly in the distinctive forms of cognitive dissonance that such cases often involve. What’s more, the person often exhibits patterns and dispositions of thought, feeling, and action sufficient to warrant taking belief in the relevant proposition to be an aspect of the person’s current view.

It would be off-key, too, to reinterpret discovery of error as just retrospective recognition that one’s former view was incorrect. When one says, reeling in stunned surprise, “My goodness, I’m wrong,” this isn’t just a report that one’s past view was wrong, nor a mere signal that one’s beliefs have suddenly changed. It often reflects the
experience of a clash between an aspect of one’s view and reality. One is speaking from within the perspective partially constituted by one’s current view, acknowledging it as one’s view and also as incorrect.¹⁸

One might be tempted to explain away reasonable double-checking as just suspension of judgment, ambivalence, or compartmentalization. However, the phenomenon has a more complicated structure than suspension of judgment, insofar as the person already believes that p. Nor is it simply a matter of ambivalence, since the person is engaged in the critical scrutiny of one of her own beliefs. And it would be a mistake to try to understand the phenomenon in terms of “compartmentalization” in any thorough-going way. In some cases of double-checking the person might say “I am of two minds,” but it is crucial that there is a single “I” whose perspective is that of being “of two minds”. Otherwise the case is not one of reasonable double-checking but rather some form of dissociation.

Further support for taking these phenomena at face value comes from cases of pathology. One evening, after leaving my office, I began to worry that the door wasn’t closed properly. I had a clear memory of checking the doorknob as I left, and I fully believed the door to be shut and locked. Nonetheless, I found myself thinking insistent, worried thoughts to the effect that I might have made a mistake, the door might not be shut, and someone might steal the computer equipment. Even as I struggled with those thoughts and feelings, I fully believed that the door was shut. Still, I doubted myself. I was filled with anxiety focused on myself and my capacities even as I fully believed that the door was locked.
Such cases highlight a feature of self-conscious belief that is shared with non-neurotic cases and makes them both possible: one can stand in various complicated evaluative stances in relation to one’s own self-conscious beliefs, stances that are akin to doubt but apply in the first instance to oneself and one’s attitudes. These are not aptly modeled simply as a reduction in credence, since that sort of approach cannot capture the significant internal structure involved in these complex attitudinal situations. There is room for a question of self-trust here that does not undermine one’s ability to relate to the belief as part of one’s own perspective. To do justice to the person’s total perspective in such cases, it seems that we have to say precisely that the person believes that p but neurotically lacks confidence in herself, is (reasonably) double-checking, or recognizes that her belief is false.

D. The Upshot

In the moment when I discover that I am in error or engage in reasonable double-checking, I do not endorse my belief as true, as supported by the evidence, or as meeting some particular epistemic standard. But I do occupy the subjective standpoint of the belief. Self-conscious belief consequently does not constitutively require endorsement of the belief. Moreover, when one reasonably double-checks one is not being irrational, so self-conscious belief does not even rationally commit one to endorsing one’s belief as true, well-supported by the evidence, or in accord with some other epistemic standard. Someone who fails to regard her self-conscious belief in one of these ways is not thereby and in that very respect necessarily vulnerable to rational criticism.
There is likewise no particular tie between self-conscious belief and Evans’s “Transparency Procedure.” In the moment in which I recognize that my belief is false, I will not be able to self-ascribe the belief by considering the relevant evidence regarding p, nor will I take myself to be able to do so. In fact, Evans’ claim, that to correctly self-ascribe belief regarding p “I must attend … to precisely the same outward phenomena as I would attend to if I were answering the question” whether p (Varieties, 225), is simply false. If, in the moment of discovering my belief is in error, I were to self-ascribe just on the basis of consideration of whether p, I would miss something important about my self-conscious view at that moment: namely, that I believe that p. In order to see the clash, and that I have a view that is in error, I must self-ascribe on some other basis.

Moreover, self-conscious belief doesn’t rationally commit one to being able to self-ascribe one’s belief via Evans’s Transparency Procedure. Someone who is double-checking might reasonably think, “I believe that p, but am I able to self-ascribe this belief by considering the evidence for and against p? That depends on whether I am currently right about what the evidence supports, and that’s precisely what I am checking.” She could be reasonable in both believing that p and wondering whether she has read the evidence in her possession correctly, and in that case she would not be rationally committed to taking herself to be able to self-ascribe her belief that p even by considering the evidence currently in her possession. The Transparency Condition can thus fail even in conditions of unimpaired rationality.

Similar points apply to the suggestion that self-conscious belief involves treating the question as settled. Even while self-consciously believing that it is raining, a person might find that she has reason to double-check whether she is right.
In sum, it seems that rational endorsement is neither a constitutive necessary feature nor a normative requirement of the stance involved in self-conscious belief. One can first-personally treat a belief as part of one’s own view, self-ascribing the belief from the subjective standpoint it partially constitutes, without any requirement that one endorse it as true, well-supported, or in accordance with some other epistemic standard, and without being able – or taking oneself to be able – to self-ascribe it by considering the facts or reasons that bear on it.

6. **Elucidatory Interlude: Objections and Replies**

It might be objected that what I have been saying is unintelligible. To believe a proposition is, in some sense, to be committed to the truth of the proposition; that’s part of the concept of belief as it applies to mature believers. This unexceptionable point might be thought to show that there is no gap between self-attributing a belief from the subject position and committing yourself to endorsing your belief as true (Stroud, 2014). However, in avowing that it is part of your subjective perspective that p is true, you don’t thereby endorse that aspect of your subjective perspective, nor need you even commit to endorsing it. As noted before, having a certain commitment doesn’t entail being committed to endorsing that commitment. It might also be part of your subjective perspective that you leave open whether your attitude towards p is correct.

Still, it might be wondered how a rational thinker could avoid endorsing her self-conscious belief upon reflection. If I self-consciously believe that p, then it is part of my current perspective both that p is true and that I believe that p. How, then, could I not be
rationally committed to the obvious conclusion if I consider whether my belief is true and don’t lose my belief in the process?

Here’s how. A thinker need not endorse using p as a premise simply in virtue of self-consciously believing p to be true, as cases of reasonable double-checking bring out. Self-conscious belief involves a disposition to treat p as a premise, but, again, this disposition is defeasible. Someone might quite reasonably self-consciously believe p, reasonably refuse (for the time being) to endorse her belief that p as true, and thus reasonably refuse (for the time being) to reason from p or use it as a premise in inference. One can thus self-consciously believe that p while rationally refusing to draw the conclusion that one’s belief is true and without rationally committing oneself to that conclusion.

Admittedly, there is rational tension if a person continues to self-consciously believe that p even while recognizing that p is false. But this doesn’t show that a rational commitment to endorsing the belief as true is intrinsically involved in self-conscious belief. After all, I can – without rational tension – self-consciously believe that p even while double-checking whether my belief is true. And the rational tension that results when I do recognize that not-p doesn’t depend on my being rationally committed to endorsing my belief that p as true; it is generated merely by requirements pertaining to the attitudes themselves (e.g., don’t also believe that p when you recognize that not-p).

Still, it might be argued that self-conscious belief rationally commits one at least to regarding one’s belief as epistemically permissible. Here, however, discovery of error provides an important test case. When a person recognizes that she believes p and is wrong, her position (considered just as a time-slice) is rationally problematic in various
ways. She both believes that p and recognizes that p is false. Moreover, she both believes that p and, insofar as she is attempting to follow the relevant norms, takes it that she should not believe p (since p is false). However, the proposal before us is that her situation is *triply* problematic because she both takes it that she should not believe p (because p is false) and is *also* rationally committed – merely in virtue of self-consciously believing that p – to the claim that it is epistemically acceptable for her to believe that p. I see no reason to think this additional rational incoherence must be involved. We have already captured what is problematic about her situation when we list the first and second points.

This verdict is supported by the parallel with other norm-governed attitudes. If I am angry with you and admit that my anger does not satisfy the relevant norms, my *anger* may be irrational, and I may be rationally required to give it up. But I am not also thereby guilty of the additional irrationality of being committed to the acceptability of my anger even as I admit its unacceptability. And this is so even if I aim to accord with the relevant norms. I can say, without the rational incoherence that would appear if I were also committed to regarding my anger as acceptable, “I’m very angry with you, even though I now see that I have no good reason to be angry. Give me some time to cool off before we talk about this more.” Self-consciously holding an attitude while aiming to accord with norms governing that attitude need not rationally commit one to endorsing the attitude as acceptable.

Certain paradigms of psychodynamically-generated irrational belief are illuminating in this regard. A man might find himself feeling guilty when kindnesses or favors are done for him, feel that he is in danger of being caught taking something he
doesn’t deserve if he accepts positive offers from others, experience his legitimate and unproblematic requests for help as manipulative maneuvers robbing others of their riches, find himself believing that accolades that come his way are the gains of trickery and deceit, feel terror that if others knew the truth they would hate and cast him out, and the like. With a great deal of therapeutic work he might gradually come to see that all of these responses are unified by a belief that he is a thief, a belief that is a fundamental aspect of his current subjective perspective on the world even though it is manifestly false, has no basis in fact or support from any evidence, and is held in place by irrational, largely unconscious, psychodynamic processes. He might gradually come to be able to self-ascribe this belief from the subject position, so that he can say, from the subjective position of the belief itself, “I believe that I am a thief.” However, in light of his recognition that this belief is entirely a tissue of fantasy, he might quite appropriately decline to assert outright “I am a thief” or to reason from this claim as a premise. He might seek to minimize the impact of this belief on his emotional life and interactions with others.

Such a person does not endorse this belief as meeting any of the norms governing appropriate believing, nor is he rationally committed to doing so. His belief is irrational, and insofar as he has it he is functioning irrationally, but there is not some further irrationality that arises from his relating to his belief fully self-consciously while not endorsing the belief as epistemically appropriate. Nothing in his subjective position rationally commits him to regarding it as epistemically appropriate to believe as he does, nor is he rationally criticizable for failing to regard this belief as epistemically acceptable. In fact, his subjective position rationally commits him to endorsing the claim that he
ought not believe that he is a thief. That’s part of the sadness and difficulty of his situation.27

These sorts of pathological cases highlight - by being parasitical upon - the features of ordinary self-conscious belief that make double-checking and discovery of error not only possible, but normatively acceptable in the particular ways that they are.

In fact, if it weren’t for our capacity to self-consciously hold a belief without endorsing it, we could not become conscious of these pathological beliefs from the subject position without falling into psychosis. This capacity is thus fundamental to our ability deal rationally with certain aspects of our irrationality.

7. Changing One’s Mind

Though cases of irrationality provide vivid examples of self-conscious belief without endorsement, the phenomena that I have been stressing are not an odd sideshow to our activities of deliberation, reasoning, and inference. They highlight a fundamental feature of life with epistemic norms. To be able to self-consciously deliberate in a way that is guided by norms, one must be able to consider whether one’s current view satisfies the relevant norms, and one must be able to recognize on occasion that it doesn’t. Active self-conscious guidance by epistemic norms thus requires one to be able to take the measure of the distance between where one is and what the norm requires. This requires that even while occupying one’s view as one’s own view, one must be able to recognize that the evidence does not support the truth of one’s belief or even shows it to be false. The capacity to occupy states of mind that, when considered as time-slices, look like
paradigms of irrationality is thus crucial for self-consciously reconsidering one’s position, recognizing error, and changing one’s mind.

The mere recognition that one is fallible is no substitute for this capacity. Try to imagine a creature that recognizes its own fallibility and is capable of fully self-conscious rational deliberation, but lacks the capacity I’ve been emphasizing. Such a creature could recognize the conceptual gap between belief and truth. But if it asked itself, “Is my current belief that p true?”, its answer would be, trivially, “Yes”. And if it reconsidered the evidence bearing on p and found that not-p, it would be incapable of recognizing that its own view is incorrect. It could only report that it formerly incorrectly thought p but now thinks not–p. It could never consciously change its mind with full reflective transparency or engage in fundamental forms of critical reconsideration that are crucial for rational deliberation.

To see this, consider what would be involved in such a creature’s self-consciously changing its mind. Suppose such a creature currently self-consciously believes that p. It now finds some evidence against the truth of p. How is this creature to answer the question of whether it should change its view? The only option here would be to cease believing that p and to deliberate afresh. After all, the creature cannot self-consciously continue believing that p even while asking (as a genuinely open question) whether its belief that p is true or supported by the evidence; that’s precisely the capacity it lacks. However, giving up the belief and deliberating afresh is not necessarily a rational option for the creature either, since one can’t acceptably just stop believing on a whim. There are epistemic norms governing these things, and so whether it is appropriate to stop or continue believing that p depends on whether certain criteria are satisfied: Does the
change in the evidential situation warrant ceasing to believe that p? To answer that question, our creature has to consider whether its belief that p meets the relevant criteria. But in order to do that, our creature has to self-attribute belief that p, as an aspect of its current view, and consider as an open question whether it is appropriate for it to believe that p given its current evidential situation. And that is precisely what it cannot do. This isn’t a question that our creature can answer by deliberating afresh about p, for the question before it (as a precursor to deliberating afresh) is precisely whether its current position is such that it should continue or cease believing that p. So the only strategy that would appear available won’t in fact work. We thus cannot really imagine a creature that is both capable of fully self-conscious rational deliberation and lacks the capacity that I am emphasizing. What we end up imagining instead is a creature incapable of self-conscious rational change of mind.\textsuperscript{28}

Of course, it is possible that such a creature might engage in conscious \textit{first-order} rational deliberation. It might ask whether p, treating this as an open question, consider evidence for and against p, and then conclude that not-p. But such a procedure would not be fully self-conscious rational deliberation, for it would not include the creature’s keeping tabs on its own belief-state throughout the process. To do that, the creature must have in view what its belief-state is from the beginning – when it believes that p – through the transitional process during which it considers the evidence and changes its mind. At some stage in this process the creature must regard itself as believing that p while holding open the truth or rational acceptability of its belief. That’s what our creature cannot do.
The capacities involved in “transparency” will not suffice to fill the gap. It might seem that when the creature is inquiring into whether continued belief is appropriate, its answer to the question, “Whether p?”, must be “p,” so that it could self-attribute belief that p by appending “I believe…” to its answer to this question. However, this suggestion neglects the broader normative context. Given the evidential situation, the creature quite appropriately might not be prepared to answer "Yes, p" in response to the question "Whether p?", just as it might quite reasonably decline to utilize p as a premise. So if it is to self-ascribe belief in the suggested way at the crucial stage in the deliberative process, the creature must proceed in a way that ignores its own understanding of the normative situation. That is not fully self-conscious rational deliberation. At this stage in the process, fully self-conscious rational deliberation will require a form of self-attribution other than transparency.29

One might try to evade the argument in other ways. Suppose that we add that our creature not only recognizes the fallibility of its belief about p, but also the fallibility of its stand on that belief’s rational acceptability. That’s still not enough to enable fully self-conscious rational deliberation. Recognizing the fallibility of its stand on the belief’s rational acceptability isn’t yet the same thing as treating this as a genuinely open question, and to treat it as an open question for genuine inquiry would require, one level up, the very capacity that our creature lacks. We might suppose that the creature instead engages in a kind of compartmentalization, simultaneously taking a stand in one compartment while considering the appropriateness of that stand in another. But in order for this to get anywhere, there also has to be the possibility of self-conscious communication across the compartments, bringing them into a single self-conscious
perspective. That too requires a capacity our creature lacks, since it cannot bring together, in a single moment of self-consciousness, the self-conscious belief that \( p \) \textit{and} recognition of that belief’s unacceptability. Finally, we might suppose that the creature engages in an “as if” exercise of considering whether the evidence rationally supports believing \( p \). But the creature couldn’t take the upshot of this pretense to have any significance for its continued belief that \( p \), since (by hypothesis) it could never recognize that while it believes that \( p \), its belief that \( p \) is not rationally acceptable. (It might be asked why the belief couldn’t simply “go away” at the moment when the two compartments are brought together, or when the “as if” exercise is concluded. This might happen, but it wouldn’t be an exercise of fully self-conscious, deliberative change of mind. The creature would merely witness a change in its beliefs.)

It might be suggested that without ceasing to believe that \( p \) the creature could “suspend” its belief while genuinely inquiring into the rational acceptability of believing that \( p \). However, all this would mean -- if not that the creature engages in the sort of compartmentalization or “as if” exercise described above -- is that the creature self-consciously believes that \( p \) but does not make use of its belief while engaging in a genuine inquiry into the rational acceptability of so believing. And this would get our creature nowhere, unless it is prepared to recognize that its current view is rationally unsupportable. This is just another way of describing deployment of the ability at issue.

In fact, without this ability, the creature’s recognition of its own fallibility wouldn’t come to much at all. When we say to people, ‘You have to consider the possibility that you might be wrong,’ we aren’t just asking them to acknowledge a merely notional possibility, only retrospectively recognizable as actualized. Rather, we are
asking that they recognize the current rational relevance of the possibility of error, and we suppose them capable of exploring this possibility in their self-conscious considerations about what to believe. So, for instance, we suppose that they have the capacity to reconsider their reasons in order to determine whether their belief is indeed as well-supported as they take it to be. Someone who lacked this capacity would not in any robust sense grasp her fallibility as having any bearing on her rational deliberations; her admission of fallibility would be merely an anemic admission that it is possible that her view might change in such a way that she can later see that she was previously wrong.

The capacity for self-conscious reflective scrutiny requires more: the ability to recognize one’s view as one’s own view even while holding in abeyance the question of endorsement. Full recognition of the deliberative significance of one’s fallibility thus implicates precisely this ability.

A physicist, for example, might attempt to reproduce a well-established result (call it “R”) by using a more sensitive apparatus. She might fully believe R and believe that her belief is true, justified, and indeed that she knows that R. Her recognition of human fallibility and the aspirations of science might nonetheless make it reasonable for her to double-check R in this way. Suppose, then, that possibly disconfirming evidence begins to pile up. She might unreflectively respond by losing her belief that R. But she might instead engage in a process of self-conscious deliberation, asking herself at various points whether the evidential situation warrants revising her view. This might be a difficult, non-trivial question, requiring her to consider a wide range of complex issues. To ask this question, she must self-consciously hold her view even while taking it to be a genuine question whether her view is warranted by her evidence. This is a question
which is not made possible merely by her recognition of her own fallibility, not even by acknowledgement of her fallibility about whether her evidence warrants believing R. But at the same time, these further capacities are implicated in the recognition of fallibility with which the whole experiment began. After all, the project was undertaken precisely with an eye to the possibility that it might eventuate in a fully self-conscious deliberative revision of her view.

8. Envoi: The Bigger Picture

I have argued that fully self-conscious adherence to epistemic norms requires the ability to self-consciously hold a belief without endorsing it as true, as well-supported by the evidence, or as meeting some other epistemic standard. As detailed in section 1, this result has significant implications for a number of broad philosophical agendas. These implications are not simply negative. The upshot is a different framework and set of starting points for philosophical theorizing in these areas.

First, there are consequences for our understanding of belief itself. For the reasons I have explored, we need a notion of belief that plays the role I’ve marked out. An adequate account must contribute to our understanding of what is going on in self-conscious critical reconsideration of one’s own beliefs. It must help us understand what is transpiring when someone self-consciously discovers that her view is wrong, and also how what goes on in that case is connected with ordinary self-ascriptions of belief from the subject position. It must help us understand the possibility of self-conscious deliberative change of mind.30
These same phenomena likewise require that we place the notion of belief within an understanding of a person’s conscious subjective perspective as something that neither needs to be, nor needs to be taken by the person to be, rationally unified. A person’s subjective perspective can simultaneously include a commitment to the truth of p and a genuine concern that the commitment may not be correct. It can contain both a commitment to the truth of p and recognition that p is false. It can be even more complicatedly fractured, as in certain cases of motivated irrationality. All this can then be part of the person's single self-conscious perspective.

As with particular self-conscious attitudes, so too with regard to her recognition of this single subjective perspective as her own. Here too endorsement need not play any central role, nor need the thought that the belief is part of an overall rational unity. Think about a person who has two incompatible loves. Both are orientations of the whole person: his total orientation involves loving – with all the dispositions involved – two things in such a way that incompatible requirements or motivational tendencies are unavoidable. He need not endorse that position, nor regard it as a rational unity, in order to self-consciously live from it. A similar point applies to belief. Sometimes one’s overall position is just a mess. One need not endorse the mess in order to recognize it as one’s own, nor need one take it to be any more rationally unified than it is.

Fortunately, one can often bring some order to the mess. Certain aspects of a person's total perspective are sometimes privileged over others. When someone recognizes that her belief is in error, she occupies a complex subjective perspective that integrates in one self-consciousness both the subjective perspective of someone who believes that p and also a broader subjective perspective in which it is recognized that this
belief is incorrect. She thus occupies a complex overall subjective perspective in which she takes various stances towards other aspects of that very perspective. This moment involves a kind of rational order, but not one of simple consistency.

Here, in rough outline, is what such integration might look like in a case of recalcitrant false belief. Someone might have the relevant feelings of conviction and the dispositions and patterns of response involved in believing that \( p \) -- the ‘fact’ of \( p \) might be part of her subjective world -- and yet because the weight of reason indicates the falsehood of her belief, she might refuse to assert, reason from, or act on \( p \). In this kind of case, she attempts, in light of the other things she believes, to quarantine her belief that \( p \) in relation to her self-conscious reasoning and action. This further response, along with the subjective perspective involved in her believing \( p \) and in her believing the other things she believes, thus forms a total complex perspective in which she believes that \( p \) while not endorsing that belief. There will be analogous moments during self-conscious processes of rational reconsideration and deliberative change of mind.

This sketch raises a host of interesting philosophical issues. But as we consider them, we must keep clearly in view a fundamental lesson about how we can best understand ourselves as reasoning creatures. Self-conscious rational functioning, both in emotion and belief, crucially involves the ability to get a certain sort of view on one’s attitude even while occupying the subjective position of the attitude. That can often involve simultaneously occupying two different, incompatible viewpoints incorporated into a single subjective perspective. In fact, negotiation of incoherence and conflict is essential if there is to be even the possibility of self-conscious, active change of mind. The capacity for fully self-conscious rational agency therefore rests upon and crucially
involves the very psychological capacities that enable fundamental irrationality. This is not the banal point that the possibility of irrationality is a standing danger for a rational being. Rather, our capacity for fully self-conscious rational self-governance makes use of and depends upon the capacity to experience a rationally incoherent total perspective as one’s own.32

Bibliography


Byrne, Alex. 2005. ‘Introspection’, *Philosophical Topics*, 33, pp. 79–104.


*The First-Person Perspective and Other Essays*, Cambridge University Press.


Siegel, Ethan. “The Power of Admitting ‘I’m Wrong’,” posted April 17, 2012, at: 

http://scienceblogs.com/startswithabang/2012/04/17/the-power-of-admitting-im-
wron/ (accessed October 27, 2014).


*Introspection and Consciousness*, Oxford University Press.


Publishing.

---

1 Two very different approaches in this tradition are Korsgaard (2009) and Frankfurt (1988).

2 Some philosophers seem to think that all that is really going on here (to put the view 
archly) is that we call evidence to mind and then watch as our beliefs revise in response. 
A subtler opposing view holds that while there is first-order reasoning, there is no form of 
thinking which “involves conceiving of yourself as its agent and is a method by which 
you can directly control your own beliefs” (Owens 2011; see also Owens 2000).

3 Shoemaker (1968, p. 8) influentially links immunity to error through misidentification 
to the kind of information sources upon which the relevant judgments are based. The 
theme is echoed and developed in Evans (1982) section 7.2.
This is, I think, the most minimal way to characterize the important lesson of section 1.4 of Moran (2001).

Burge (1998) stresses similar points, though he puts them to very different use.

Similar considerations are suggested by Shoemaker (2012, p. 244, and 1996).

To speak of “rational commitment” here is to suggest that one would be open to rational criticism if, and insofar as, one failed to regard one’s belief in this way.

Moran (2001, 2012) appears to regard the Transparency Condition as articulating a normative demand concerning the first-person stance. However, the literature on Transparency has instead focused primarily on its relation to self-knowledge. Significant discussions include Shah and Velleman (2005), Fernandez (2005), Byrne (2005, 2011), Silins (2012), Gertler (2011). Byrne’s inferentialist account carries a commitment similar to the one highlighted above, insofar as the inference he describes precludes our ever self-consciously believing something which we take to be false. (My characterization of the Transparency Procedure in the main text above is neutral between Byrne’s inferentialist approach and non-inferential interpretations.)

For instance, in certain situations of emotional arousal many patients with Borderline Personality Disorder cannot properly consider in situ the possibility of a mismatch between their beliefs and the world. A stereotypical conversation between a psychotherapist and a BPD patient might run as follows. Therapist: “You think I am like your boss and constantly criticizing you.” Patient (with rising anger): “Of course I think you are like my boss. You are always criticizing me; you are doing it right now.” The BPD patient cannot critically scrutinize his belief that the therapist is criticizing him, because the patient cannot get the right sort of perspective on his own belief.
Psychoanalysts Peter Fonagy and Mary Target characterize such functioning as “psychic equivalency mode” and identify it with certain early stages in child development. Someone functioning in psychic equivalence mode will have recognizable difficulties with self-conscious rational reflection and affect regulation. See Fonagy and Bateman (2004) and Fonagy, Target, et al. (2005).

This characterization is compatible with a broadly Davidsonian, holistic approach to belief attribution, though Davidson (1982) goes astray in supposing that rationality constraints on belief attribution require postulating, in effect, two distinct minds in cases of fundamental irrationality. If (as has been urged by Hursthouse 1991) the distinctive form of intelligibility provided by person-level folk-psychological explanation is not primarily and narrowly rational intelligibility, then the possibility opens up of attributing much more complex, conflicted belief states to the person. Schwitzgebel (2002) offers a broadly dispositional approach to belief that fits well with the perspective sketched here.

This phenomenology is distinct from what is presented as being the case, but rather concerns the manner in which it is so presented.

David Simpson, comment posted November 2013 at
http://www.ted.com/talks/kathryn_schulz_on_being_wrong?language=en (accessed October 27, 2014). Schulz’s TEDtalk relates to a series she wrote for Slate based on interviews with prominent people about their attitude towards being wrong.

For a humorous depiction of the lengths to which people will sometimes go to avoid accepting that they are in error, see Rosenberg (2015).

Several Early Modern philosophers thematized the way in which simultaneous contradictory views of the same object can give rise to one or another form of uneasiness.
See, for instance, Hume, *Treatise*, Book II.3 sect. 9-10; Spinoza, *Ethics* III prop. 35 & scholium (concerning the pain caused by contradictory feelings of love and hatred, each of which involves - and on some interpretations just is - a judgment of goodness or badness respectively). Thanks to Kate Abramson regarding these historical matters.

15 The existence of such cases is compatible with the phenomenon of “higher-order defeat”, since for all I’ve said continued belief might be rendered irrational by stronger evidence that something is amiss regarding one’s belief or its provenance. (For discussion of this phenomenon see, eg, Lasonen-Aarnio (2014.)

16 Not every case of reasonable double-checking is of this sort. In some cases the mere fact of human fallibility makes it reasonable to run through the evidence again even though one is fully committed to endorsing one’s belief as true. However, what matters now are cases in which one continues to believe as one did, and yet while one is checking one regards it as a genuinely open question whether one’s belief is true. This involves something different from merely regarding it as epistemically possible that one’s belief is false. (If one doubts that self-conscious belief could allow for anything different from that, consider the sort of pathological self-doubt described below.)

17 This point relates the use of “I believe”, where it functions as a *hedge*, and its use to express *full endorsement*. In both cases, one self-ascribes the belief from the subject position. But when one reasonably says “I believe that p” as a hedge, refusing to assert p outright or utilize it as a premise in reasoning because one suspects one might be mistaken, one expresses one’s belief through self-attribution from the subject position but simultaneously indicates that one takes there to be reason to withhold one’s endorsement of it.
Such moments are aptly captured by the “commissive” version of Moore’s paradox, “I believe that p, but p is false.” For a related point about Moore’s Paradox, see Gertler (2011, 139-140). As Gertler notes, the “commissive” form of Moore’s Paradox sometimes expresses a step towards improved rationality.

Contra Huemer’s suggestion (2008).

Shifting to talk of degrees of belief makes little difference here. The literature on “higher-order defeat” standardly assumes that one can consider how one’s first-order credence rationally ought to be revised in light of one’s assessment of the rationality or likely reliability of that credence; the discussion thus assumes that one can self-consciously have a high first-order credence even while having a significantly lower degree of belief that one’s first-order credence is appropriate. This literature thus largely takes for granted the point that I am urging.

This point relates to a broader issue stressed by Shah and Velleman (2005), namely, that following the Transparency Procedure may lead to a change in one’s beliefs. Importantly, one can self-consciously recognize – even while relating to one’s belief in a properly first-personal way – that this possibility might well occur.

Gertler (2011, 135-42) offers related reservations, though she does not emphasize that transparency can fail even in conditions of unimpaired rationality.

Stroud writes:

Someone who believes something takes something or other to be so. For anyone who understands what he is saying or thinking in ascribing a belief to himself, there is therefore nothing in need of further explanation .... He regards the belief he ascribes to himself as true. … all believers regard the
beliefs they know they hold as true. (2014, 173).

24 For this reason, Moran goes subtly astray when he comments, “What is unavoidable from the first-person perspective …, is the connection between the question about some psychological matter of fact and a commitment to something that goes beyond the psychological facts” (2001, 76). Even if the belief is or involves a commitment regarding the facts, one can relate to that belief as a part of one’s own view without endorsing or speaking from that commitment.

25 Such phenomena are illuminatingly discussed by the psychoanalyst Ron Britton (1998).

26 The acquisition of the ability to self-ascribe the belief in this way is an important step towards psychic integration and health.

27 As this case shows, someone who self-consciously believes that p without endorsing her belief need not be in violation of a wide-scope requirement of rationality in that very regard. The person with the self-conscious pathological belief discussed above is in violation of various wide-scope coherence constraints insofar as (and because) his belief does not cohere properly with other things he believes. However, given his situation, lack of endorsement of this belief is a rational response. There is no fully general wide-scope rationality constraint that says, roughly, "Don't both believe that p and also fail to endorse your belief as epistemically acceptable." (I am grateful here for questions from Ram Neta.)

28 This argument concerns the evidential propriety of the creature’s belief, and so it might seem to leave completely open the possibility that self-conscious belief requires an endorsing stance towards one’s belief so far as truth is concerned. However, imagine a
creature that (a) cannot avow its belief without endorsing it as true, but (b) can avow it without endorsing it as evidentially warranted. Such a creature could find itself in the position of saying, “I fully approve of my belief insofar as truth-value is concerned – my belief is true – but it is an entirely open question for me whether the evidential situation supports my believing as I do.” And if it found that its current belief was not evidentially warranted, then it would have to say something to the effect of, “My current belief is not evidentially acceptable, but still, I endorse it as true.” Such a creature thus falls into objectionable dogmatism: it decouples approval of its view with regard to truth from considerations about evidence.

29 Similar problems plague more refined versions of the transparency approach. Consider the view that self-attribution is transparent to consideration of the evidence for and against the truth of p. When the creature inquires into whether the evidence warrants giving up its belief and deliberating afresh, it cannot at that moment self-attribute its belief by considering what the evidence supports, for what the evidence supports is precisely what it is inquiring about. But if the creature takes itself not to believe that p at that moment, then it must take itself to have stopped believing that p without yet having determined whether there is good reason to do so – since that was precisely what it was attempting to determine. And it won’t do to declare that knowledge of its belief-state is impossible at this juncture, because that would be to grant that fully self-conscious deliberation isn’t possible for this creature after all. This version of the transparency theory consequently doesn’t show how a fully self-conscious, rational change of mind could take place here, even given that the creature can engage in conscious first-order deliberation about what the evidence supports. Inferentialist interpretations of the
transparency procedure (Byrne 2005, 2011) aren’t any better off. According to such views, self-ascription arises from inference in accordance with the rule, “If p, then believe that you believe that p.” However, given the evidential situation at this stage in the deliberation, the creature quite appropriately might not be prepared to use p as a premise. So if the creature is to self-ascribe its belief via this reasoning at this stage in the deliberation, its reasoning must not be reflectively connected with its own understanding of the normative situation. Again, that is not a conception of fully self-conscious rational deliberation. At bottom, transparency views share a fundamental failing. They regard self-attribution of one's view – and recognition of a change in one’s view – as derivative from prior and independent first-order processes. This undercuts the very agential capacities that motivate much of the appeal of the transparency approach. It is a minimal necessary condition on being a mature self-conscious agent that one be able to decide to revise one’s view precisely because one recognizes that one’s view, understood as one’s view, has a fundamental shortcoming. That is what the transparency view cannot give us, as the above problems highlight in their various ways.

Perhaps, like Gendler (2008), we should distinguish various kinds of belief-like states, for instance those states which the agent will endorse upon reflection and whose contents the agent will assert, from some of those content-bearing states which shape the agent’s behavior. However, even if the ordinary notion of belief shatters under this sort of pressure, my substantive point remains. If we want something that can do the explanatory and normative work of the notion of “a person’s view”, and if we want to allow for the possibility of fully self-conscious change of mind, then we need to make room for an
attitude with the kind of profile that I am describing. (Thanks to Ram Neta for pressing me on this issue.)

31 Analogous imaginative capacities are stressed by Robert M. Gordon (2007), as when one engages in the complex act of imagining what it would be like first-personally to believe something that is, in the imagining, false.

32 This paper draws upon material presented at the 2013 Eastern American Philosophical Association Meeting, the 2013 Auburn University Conference on Theoretical Agency, and the 2014 University of London Institute of Philosophy and London Institute of Psychoanalysis joint conference, “Philosophy and Psychoanalysis in Dialogue”. I am grateful to the audiences on these occasions. Thanks to Matt Boyle, Gary Ebbs, Brie Gertler, Mark Kaplan, Hilary Kornblith, Kirk Ludwig, Dick Moran, and Ram Neta for conversations and comments relating to this paper. Thanks especially to Katy Abramson for helping me work out these ideas and bring them into publishable form. Research on this paper was generously supported by an Andrew W. Mellon Foundation New Directions Fellowship.