Transparency and Reflection
Matthew Boyle, Harvard University

—Work very much in progress. Please don’t cite without permission.—

"[T]he mode of existence of consciousness is to be conscious of itself... [But] this consciousness of consciousness—except in the case of reflective consciousness, on which we shall dwell shortly—is not positional, which is to say that consciousness is not for itself its own object. Its object is by nature outside it."

Sartre, The Transcendence of the Ego, pp. 40-1/23-4

1. Transparency as phenomenon and as problem

In an influential discussion of the basis of first-person belief ascriptions, Gareth Evans pointed out that we are normally in a position to ascribe beliefs to ourselves, not by seeking evidence concerning our own psychological states, but by looking to the realm of non-psychological facts:

In making a self-ascription of belief, one’s eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me "Do you think there is going to be a third world war?", I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?" I get myself in a position to answer the question whether I believe that p by putting into operation whatever procedure I have for answering the question whether p. (Evans 1982, p. 225)

Evans also held that a similar point applies in the case of first-person ascriptions of perceptual appearances:

[The] internal state [produced by perception] cannot in any sense become an object to [the subject]... However, a subject can gain knowledge of his internal informational states in a very simple way... He goes through exactly the same procedure as he would go through if he were trying to make a judgment about how it is at this place now, but excluding any knowledge he has of an extraneous kind. (That is, he seeks to determine what he would judge if he did not have such extraneous information.) The result will necessarily be closely correlated with the content of the informational state which he is in at that time. Now he may prefix this result with the operator ‘It [perceptually] seems to me as though ...’ (Ibid., pp. 227-228)

The phenomenon Evans describes here has come to be known as the “transparency” of

1 I give page references to works by Sartre first in the standard English translation and then in a standard French edition (see the bibliography for details). In various places I have modified the translation.
(certain forms of) self-knowledge. Our knowledge of our own mental states is said to be “transparent” inasmuch as we can answer questions about these states by attending in the right way, not to anything “inner” or psychological, but to aspects of the world at large.²

Although Evans discusses only our knowledge of our own beliefs and perceptual appearances, there is reason to think that the point he noticed has wider application. For instance, the mind-focused question whether I intend to ϕ is arguably normally transparent for me to the world-focused question whether I will ϕ (when the latter question is answered subject to certain restrictions),³ and the mind-focused question whether I want X is arguably normally transparent for me to the world-focused question whether it would be desirable for me to have X (again, answered subject to certain restrictions).⁴ Each of these formulations is rough and incomplete, and the task of making them sharp and complete would be little easier than, and closely related to, the task of giving a philosophical account of intention and desire. But while there is room for dispute about how to characterize specific relations of transparency, it is widely accepted that there is a significant phenomenon here for which an adequate theory of self-knowledge must account. The general phenomenon is this: I seem to be able to know facts about my own mind simply by considering aspects of the world on which my mind is directed. To obtain this knowledge, I look, as it were, not inward but outward.⁵

Evans seems to have regarded his observations as demystifying our capacity for certain kinds of privileged self-knowledge. It can seem mysterious how we are normally able to say what we believe without observing ourselves, even though we must observe another person to determine what she believes; and this can seem all the more puzzling when we note that, although each of us normally speaks on the topic of his or her own beliefs without any evidential basis, our self-ascriptions of belief are treated as authoritative in a way that the ascriptions of others who do attend to the evidence are not. But, Evans²

---

² Some philosophers seek to characterize the notion of transparency in a way that does not prejudge the question whether the relevant self-ascriptions express knowledge, but in the present essay I will assume that these ways of answering questions are ways of knowing what one believes and how things perceptually appear to one. That is certainly how we seem to treat them in ordinary life, and I will presume this naïve classification to be innocent until proven guilty. If there are grounds for denying that such self-ascriptions express knowledge, I think all the substantive points in this essay could be reformulated (more cumbersomely) to accommodate the point.


⁵ That we can acquire self-knowledge in some such outward-looking way is widely but not universally accepted. For dissent, see for instance Gertler 2011.
suggests, this should not seem strange once we recognize that a person can answer the question whether she believes \( p \) by “putting into operation whatever procedure [she has] for answering the question whether \( p \)”. For where \( p \) is a proposition about the non-mental world, it is no surprise that a subject can answer the question whether \( p \) without needing evidence about her own psychology. And if by answering this question she is automatically in a position to answer the question whether she believes that \( p \), then it makes sense that she should be an authority on this issue whenever she is capable of answering the question whether \( p \). “If a judging subject applies this procedure”, writes Evans, “then necessarily he will gain knowledge of one of his own mental states: even the most determined sceptic cannot find here a gap in which to insert his knife” (Evans 1982, p. 225). He would presumably say something more cautious about the case of perception, but the point of principle is the same: here too, Evans’s procedure is a way of getting into a position, without self-observation, to speak authoritatively about one’s own state of mind.

Many subsequent writers, however, have thought of Evans’s observations, not as providing the solution to a puzzle, but as presenting a puzzle in their own right. After all, where \( p \) is a proposition about the non-mental world, it is generally possible for it to be the case that \( p \) although I do not believe that \( p \). Indeed, this is surely my actual situation for a tremendous number of values of \( p \): I am very far from omniscient. How then can it be reasonable for me to treat the question whether I believe \( p \) as tantamount to the question whether \( p \), when the proposition that \( p \) does not entail, and does not seem even to provide defeasible support for, the proposition that I believe \( p \)? And similarly, in the case of perception, it is surely often the case that there is something in my vicinity although I do not perceive that thing. I am very far from omnipercipient, even in respect of my local environment. So again, how can it be reasonable for me to treat the question whether I perceive a certain thing as tantamount to the question whether there is such a thing “at this place now”? What justifies me in answering a question about my own psychology by looking to a seemingly quite independent fact about the world?\(^6\)

It will not suffice, as an answer to these questions, simply to point out that the contents of our beliefs and perceptions are systematically related to the contents we would express in answering corresponding world-oriented questions. This is certainly true, and it is part of what gives Evans’s observations their initial plausibility, but it does not address

---

\(^6\)This question has been pressed by André Gallois (1996), Richard Moran (2001), and Alex Byrne (2005, 2011, 2012), among others.
the heart of the puzzle. For, as Fred Dretske has emphasized in several recent papers (Dretske 2003, 2012), it is one thing to know what you think (believe, perceive, want, intend, etc.) and another thing to know that you think it. If I reach the conclusion that there will be a third world war, or judge on the basis of perception that there is a table strewn with papers in front of me, I thereby gain some information about what I believe or (apparently) perceive: I become cognizant of an apparent fact that is, as it happens, the content of a certain belief I hold or a certain perceptual experience I am having. But the procedures Evans outlines instruct me to lay claim to a further item of knowledge: the knowledge that I believe that there will be a third world war or that it perceptually appears to me that there is a table strewn with papers in front of me. It is this psychological knowledge whose warrant is in question, and the puzzle is that nothing in my apparent basis seems to supply a ground for it.

Several of the most influential recent approaches to self-knowledge take their departure from this problem. It has, in the last decade, become a principal focus of discussion, succeeding and to some extent supplanting more generic debates about the basis of “first person authority”, whether privileged self-knowledge is compatible with externalism about mental content, and so on. I believe this is a positive development, not because those earlier debates were misguided, but because focusing on the problem of transparency sharpens our sense of a crucial issue at stake in them. What is at stake is not merely how we know a certain special range of facts, but how this special variety of knowledge is related to our capacity for cognition of the wider world. Confronting the problem of transparency forces us to confront the question how our capacity to engage, theoretically and practically, with the non-mental world contributes to, or is conditioned by, our capacity to know our own minds. A number of prominent approaches to the problem can, I think, be understood as expressing different attitudes toward this fundamental question.

My aim in this essay is to survey a few of these approaches, bringing out how they bear on the question of the relation between knowledge of mind and knowledge of world, and considering their strengths and weaknesses in a way that will, I hope, clarify the motivation for the approach that I myself endorse, an approach I have elsewhere called “reflectivism” (Boyle 2011). My survey will not exhaust the actual approaches on offer, much less the ones that are possible, but those I will consider are among the most influential, and each seems to me very natural and attractive. I will argue that these views
seem as attractive as they do because each contains some element of the truth, though not the whole truth; and I will suggest that it is possible to conserve the best insights of each without excluding those of the others. This will lead me to restate and clarify the reflectivist approach, and to correct some of my earlier claims about it, in a way that draws on some suggestive ideas from Sartre.

2. Moran on transparency and deliberative agency

The prominence of the term “transparency” as a label for the relation between questions about one’s own mind and questions about the world is primarily due to Richard Moran’s work, and his account of why the question whether I believe that $p$ is (normally) transparent to the question whether $p$ (henceforth, “doxastic transparency”) will provide us with a useful starting point, both because it is grounded in certain very compelling observations about the character of transparent self-knowledge, and because the other approaches we will consider are framed in significant part as responses to Moran’s position.

Moran famously holds that doxastic transparency is explained by the fact that we normally answer the question whether $p$ by exercising our capacity to determine whether to believe that $p$ – our capacity to “make up our minds”. It is not the case, he observes, that we can always determine whether we believe $p$ by answering the question whether $p$: there are occasions on which we must discover our actual belief as to whether $p$ by observing ourselves, much as a spectator would. These are, however, pathological situations in which we are “alienated” or “estranged” from our own beliefs, in the sense that we cannot regard these beliefs as governed by our conscious assessment of grounds for taking the relevant propositions to be true. When we are not thus alienated, Moran suggests, we are in a position to know whether we believe $p$ by considering whether $p$ precisely because such consideration settles (i.e., makes it the case) that we believe (or do not believe) $p$. It is this that resolves the puzzle of doxastic transparency:

> What right have I to think that my reflection on the reasons in favor of $p$ (which is one subject-matter) has anything to do with the question of what my actual belief about $p$ is (which is quite a different subject-matter)? Without a reply to this challenge, I don’t have any right to answer the question that asks what my belief [about, e.g., whether it will rain] is by reflection on the reasons in favor of an answer concerning the state of the weather. And then my thought at this point is: I would have a right to assume that my reflection on the reasons in favor of rain provided me with an answer to the question of what my belief about the rain is, if I could assume that what my belief here is was something determined by the conclusion of my reflection on those reasons. (Moran 2003, p. 405)
Here, Moran claims only that doxastic transparency would be intelligible if the conclusion of my reflection on whether to believe p determined whether I believe p, but elsewhere he adds that such transparency is intelligible only if this is assumed (cf. Moran 2001, pp. 66-7). For, he argues, it is reasonable for me to treat the question whether I believe that p as transparent to the question whether p only if I am entitled to assume that what I conclude about whether p determines what I actually believe concerning p, and to assume this is in effect to assume that my belief is “up to me”, in the sense that what I believe concerning p depends on whether I accept that p is credible. He summarizes this idea by saying that I am in a position to have transparent knowledge of my own beliefs just insofar as I am entitled to address the question whether I believe that p, not as a “theoretical question” about what is (perhaps unbeknownst to me) the case with me, but as a “deliberative question” about whether to believe p (cf. Moran 2001, pp. 58, 63).

It is hard to read Moran’s work on doxastic transparency without feeling that he has put his finger on something crucial, but on closer scrutiny, it is difficult to say just what the insight is and how broadly it applies. For one thing, as a number of critics have pointed out, the phenomenon of transparency is not confined to conditions that are in some sense (potentially) self-determined.\(^7\) As we noted in §1, questions about how things perceptually appear also seem to exhibit a kind of transparency to world-directed questions, as for that matter do questions about appetitive desire (e.g., whether I’m hungry can manifest itself in whether a cheeseburger looks delectable). These are surely not conditions we determine to exist on the basis of reasons: they are conditions to which the question “What is your reason for X-ing?” does not apply. This gives us grounds for doubting Moran’s claim that transparent self-knowledge would be inexplicable if we were not self-determining: for it seems that in these cases transparent self-knowledge is possible, and so presumably not inexplicable, but plainly it is not grounded in our capacity for rational agency or self-determination.

Furthermore, it is not easy to see how Moran’s idea can supply a general account even of doxastic transparency. There are, after all, many beliefs I hold, and transparently know myself to hold, without (as it seems) taking a deliberative stance toward them.\(^8\) I believe, for instance, that the former President William Howard Taft was born in my home

\(^7\) Cf. Byrne 2005, p. 85, and for related observations see Finkelstein 2001, Postscript and Bar-On 2003, Chapter 4.

state of Ohio, and doubtless I once had some basis for believing this, but I cannot for the life of me recall what it was. Did I read it on a plaque, or in a history book, or hear it from a teacher? I have no idea: at this point it just stands among the countless things I take to be true without having specific grounds for holding them true. Nevertheless, my knowledge that I believe this surely meets the Transparency Condition: I can answer the psychological question whether I believe that Taft was born in Ohio simply by referring it to the factual question whether Taft was born in Ohio. Now how could the idea that I take a “deliberative stance” toward the question whether Taft was born in Ohio explain this? It is characteristic of my attitude that I do not regard this question as open to deliberation; my view is settled. And while it may be true that I would even now answer affirmatively to the question whether to believe that Taft was born in Ohio, my reason for doing so seems to be precisely my conviction that Taft was born in Ohio. So it seems that here, my answer to the question whether to believe this does not determine whether I believe it, but rather the reverse.

Finally, and most seriously, there is a crucial unclarity in Moran’s proposal. Consider a case in which I do deliberate about whether \( p \) and conclude: Yes, \( p \). If all has gone well, I now believe that \( p \). But what accounts for the fact that I am now in a position to know that this is so? Consider my cognitive situation at the moment at which I conclude “Yes, \( p \)”: as Alex Byrne asks, “Now what?” (Byrne 2011, p. 203). Evans’s transparency procedure advises me to go on to judge that I believe \( p \), but if I thereby express knowledge, what accounts for this? Moran is not explicit on this point: my warrant for so judging is evidently supposed to be connected with the fact that I have determined that \( p \) is to be believed, but it is not clear just how this is supposed to carry me across the threshold to doxastic self-knowledge. For suppose it is right that my concluding “Yes, \( p \)” is tantamount to my concluding: \( p \) is to be believed. It remains unclear how I can justifiably move from the conclusion \( p \) is to be believed to the proposition I believe that \( p \). After all, \( p \) can certainly be the correct belief to hold on the question whether \( p \) although I don’t actually believe \( p \). Indeed, as a non-omniscient being, this is surely very often my situation. So it seems that the puzzle of transparency has not been resolved, just relocated. For what justifies me in answering a question about my own psychology on the basis of a seemingly independent fact about what is to be believed?

It is possible to imagine various accounts of my warrant for moving from the judgment that \( p \) to the judgment that I believe \( p \). One view would be that my warrant for the latter thought depends on my awareness of my own activity of judging that \( p \). Such a
view might draw inspiration from Moran’s emphasis on the importance of rational agency, and from the comparisons he draws between our transparent knowledge of our own beliefs and our agential knowledge of our own intentional actions. And such a view has in fact been defended by Christopher Peacocke (1998, 1999, 2008), who holds that our capacity for transparent knowledge of what we believe is grounded in a special kind of “action awareness” we have of what we consciously judge. A different, more austere view would be that I infer that I believe \( p \) from the (apparent) fact that \( p \) itself. A view of this sort has been defended by Alex Byrne (2005, 2011), and it has the advantage that it holds on to the thought Evans made attractive: that we can acquire knowledge of our own psychology, not by consulting some antecedently-available form of self-awareness, but simply by looking \emph{outward} toward the wider world.

I think that each of these views is natural and attractive, and that each of them contains an element of truth, but I want to argue that neither is satisfactory as it stands. That will be my project in the next two sections. Before turning to this task, let me suggest a lesson to take from Moran’s discussion. Whatever we think of the link he asserts between doxastic transparency and rational self-determination – and I myself think there is a real insight here, one that can be defended against the objections raised above\(^9\) – we should, I believe, aim to respect an observation he makes about the general \emph{character} of transparent self-knowledge. It seems to me that there is something very compelling about his idea that we can know our minds transparently only insofar as our self-knowledge is not spectatorial or theoretical in character. The point here is not merely that such knowledge is available to us without self-observation or other inference from supporting evidence; it is that our \emph{stance} toward the relevant facts is not that of a spectator.

There are some remarks in Wittgenstein’s \emph{Remarks on the Philosophy of Psychology} that help to bring out what I have in mind. Wittgenstein is considering the question why I don’t look for behavioral evidence before ascribing intentions to myself, and in particular why I don’t seek to ascertain whether I am lying by looking for the kinds of behavioral evidence I rely on in assessing whether another person is lying. He begins by presenting the response of an imagined interlocutor:

\begin{quote}
“But I know that I am lying! What need have I to draw conclusions from my tone of voice, etc.?" — But that’s not how it is. For the question is: Can I draw the same conclusions, e.g. about the future, from that ‘knowledge’; can I make the same \emph{application} of it, as of observed signs?
\end{quote}

\(^9\) For such a defense, see Boyle 2009, 2011a, 2011b.
Why don’t I make inferences from my own words to a condition from which words and actions take their rise? In the first place, I do not make inferences from my words to my probable actions.\(^{10}\)

Wittgenstein imagines his interlocutor saying, in effect: “I don’t look for behavioral evidence that I am lying because I don’t need to. Given that I lie intentionally, I know this immediately.” But, Wittgenstein replies, that is not the fundamental point; for it is not merely that I do not require evidence to ascertain this sort of fact, but that this form of “knowledge” does not play the role for me that it might for another person. I do not, for instance, normally treat my awareness of my own intention (or my self-ascription of an intention) as a basis for predicting what I will (probably) do; to do this would already be a step toward abdicating responsibility for seeing to it that I do it.\(^{11}\) More generally, when I am not in an alienated relation to my own mental states, my knowledge of my own mind does not function for me as mere information about my psychological state, on the basis of which I can speculate about what I am likely to do. When I am in a non-alienated condition, I do not merely know about my own states, but speak from the standpoint of these states in making the relevant self-ascriptions, and my capacity for transparent self-knowledge is a marker of this. This is vaguely expressed, but I think it gets at something important, something Moran’s work has done more to bring out than any other work on self-knowledge that I know of.\(^{12}\)

Note that this point can be accepted by philosophers who reject Moran’s idea that the only alternative to a “theoretical stance” toward oneself is a “deliberative stance”. It may be that Moran overestimates the closeness of the tie between not taking a theoretical or spectatorial stance toward ones own mental states and treating those states as open to deliberation. Perhaps this opposition characterizes only certain kinds of states (belief and intention?), if it applies anywhere. Be that as it may, it remains true in general that our relation to those mental states we can know transparently is not theoretical in character. Perception can again serve as an example: I can have transparent knowledge of my own

---

\(^{10}\) Wittgenstein 1980, Vol. I, §§739, 814. These remarks are part of a sequence that returns again and again to this idea, that the distinctive thing about avowals of one’s own present mental states is not merely that they are known immediately, but that the relevant knowledge plays a different sort of role in our lives from mere information. Cf. §§705, 714, 744.

\(^{11}\) Cf. Moran 2001, Chapters 3 and 5.

\(^{12}\) It is also, in another way, an important theme in work inspired by Wittgenstein’s idea that our privileged knowledge of our own minds is grounded in our capacity to express our mental states through psychological self-ascriptions. See especially Finkelstein 2001, Chapter 5.
perceptual appearances, but a person suffering from "blindsight", who is able to perform certain perception-dependent tasks when prompted but who is not conscious of perceiving, cannot know his own perceptual appearances (at least, the ones in his “blindsight” region) in this way. He is alienated from his own perceptual appearances in a way broadly analogous to the way in which a psychoanalytic patient may be alienated from one of his own beliefs: he can say what he apparently perceives only by observing his own behavior, and his authority on this topic is no greater in principle than that of another person. And like the psychoanalytic patient, he can relate to any information he does acquire about the relevant mental states only in a theoretical way, as a basis for predicting (or “retrodiciting”) how he is likely to respond in various circumstances. But our normal relation to our perceptual appearances is not like this: it is not merely that we can know how things perceptually appear with an immediacy unavailable to the blindsighted person; it is that our knowing that we are perceiving a flash of light in a certain place figures for us, not as a basis for predicting that we will point to that place when prompted, but as an immediate reason to point to that place. Our stance toward our own perceptual self-knowledge is not theoretical/spectatorial, though it is also not deliberative: this knowledge is certainly not grounded in any ability to deliberate about whether to have certain perceptual appearances. If it were, perception could not be the sort of basis for knowledge that it is.

3. Byrne on transparency and inference

Alex Byrne has done more than any other recent author to bring out the generality of the problem of transparency and to emphasize how this problem rests on the apparent need for a cognitive transition from world to mind. The difficulties I have raised for Moran's view are patterned on difficulties posed in Byrne’s work. Byrne’s conclusion from these difficulties is that our capacity to make up our minds through deliberation does not explain our capacity for transparent self-knowledge, even in the case of belief. His own solution to the problem of transparency is simple in its main idea, though nuanced in how it treats different relations of transparency.

The main idea is this. Our capacity for inference is in general a capacity to implement in our psychology certain rule-governed transitions between one content and

---

13 If he had such information, he could, of course, exploit it in a secondary way as guidance about the state of the world, but any actions he performed on this basis would not be actions performed simply on the basis of perception. Our normal actions on the basis of self-known perception remains action based simply on perception.
another content. The problem of transparency is that the transitions that lead us to self-
knowledge – for instance, the doxastic inference schema

\[
\text{BEL: } \quad p \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \qu
In \textsc{see}, “[... x ... ]v” is a “v-proposition”: a proposition ascribing to \textit{x} only properties characteristically available to vision (shape, orientation, depth, color, shading, movement, etc.). The \textsc{int} inference is supposed to be defeasible, and the subject must refrain from drawing it if he takes himself to believe that he will \phi on the basis of good evidence (which is a psychological matter, but which he can ascertain on a neutral basis via \textsc{bel}.

I have learned a great deal from Byrne’s work on these topics, but I am not satisfied with this approach to the problem of transparency. I have three objections. First, it seems to me that Byrne’s approach does not explain the rational intelligibility of the relevant transitions from the subject’s own standpoint. Our capacity \textsc{for} (personal level) inference is a capacity to make cognitive transitions in virtue of seeing one or more (seeming) truths as supplying some sort of \textit{reason} to accept some further proposition as true. The conclusions I reach through inference are not just convictions that appear unaccountably in my mind; they are convictions for which I take there to be some intelligible ground, and this is what makes these conclusions sustainable in the face of the capacity for critical scrutiny – for asking “why?” – that belongs to rational subjects as such. The problem of transparency, it seems to me, is not merely the worry that inference principles like \textsc{bel} would be unreliable, but a concern about how a subject who draws such inferences could understand them to be reasonable. I cannot see how Byrne’s account sheds light on this.

Suppose for the sake of argument that a subject concludes that he believes \textit{p} by inferring according to \textsc{bel}. Let him now ask himself on what grounds he holds it to be true that he believes \textit{p}. Citing the ostensible fact that \textit{p} looks obviously unhelpful: this by itself has no tendency to show that it is true that he or anyone believes that \textit{p}, as Byrne himself admits. What would support the subject’s conclusion, of course, is the fact that he, the maker of the inference, accepts that \textit{p}. But to represent \textit{that} as his basis for accepting that he believes \textit{p} would be to presuppose that he already knows his own mind on the matter, and that would undermine Byrne’s account. So Byrne’s proposal appears to face a dilemma: either it represents the subject as drawing an inference that he should find rationally

\begin{align*}
\text{SEE:} & \quad [\ldots x \ldots ]v \; \& \; x \text{ is an } F \\
\text{INT:} & \quad \text{I will } \phi \\
\text{I see an } F^{15} & \quad \text{I intend to } \phi^{16}
\end{align*}

\textit{Belief (personal level)}

\text{For a related but more complex proposal that draws on Byrne’s general framework for understanding transparency, see Setiya 2011.}

\textit{These remarks expand on points made in Boyle 2011b.}
unintelligible on reflection, or else it requires him to have a kind of ground that would contradict the basic idea of Byrne’s approach.

If we grant for the sake of argument that there could be a rational subject who was disposed to make cognitive transitions according to Byrne’s BEL-schema, I suppose such a subject could come to appreciate Byrne’s arguments for the claim that this disposition is reliable and safe, and then she could have a kind of second-order approval of her first-order disposition to make the BEL-transition. But – in addition to the fact that this seems to be too roundabout and sophisticated a rationale for a step that we ordinarily take to be straightforward and unproblematic – this would be a post hoc approval, not an understanding in virtue of which the subject makes the relevant transition itself. The structure of Byrne’s account requires that the basic transition be from a proposition sheerly about the world to a proposition about the subject’s own mind, and this appears to require that the subject’s disposition to make this transition must be automatic, not rational. That looks to me, not like a solution to the problem of transparency, but the biting of an unappetizing bullet.

My second objection is that reflection on familiar cases of transparency suggests that the cognitive transitions we make are not, as Byrne suggests, transitions from a sheer proposition about the world to a proposition about the subject’s mind. To see what I mean by saying that the basis is not a “sheer” proposition about the world, consider the transition from I will φ to I intend to φ. I think Byrne is right that when I take it that I will φ, on a certain sort of basis, this also warrants me in judging that I intend to φ. This would constitute a vindication for Byrne’s approach, however, only if my basis for judging that I intend to φ were neutral, in the sense that its availability did not presuppose an awareness of my own intentions. Now, the proposition I will φ is superficially neutral: it does not explicitly refer to the subject’s present mental state, but only to what he will do at some point in the future. But if we think carefully about the kinds of circumstances in which I might, on the basis of thinking I will φ, be warranted in thinking I intend to φ, we will see that a quite special use of “will” must be at issue here.

Let us stipulate the existence of an “intention-based” sense of “will” (“willi”), whose use in joining a subject with an activity-verb implies that the subject will perform the relevant activity in virtue of a present intention to do so.¹⁸ We can distinguish “willi” from a

¹⁸In English, it is natural to use “I am going to” in a way that carries this implication: when a person says “I am going to φ”, this normally (though not universally) implies that her future φ-ing is settled in
“will” of blank futurity (“will_{BF}”), which merely asserts that the subject will at some future
time engage in a certain activity, leaving it open what makes this the case. In the “will”-
- sense, it is true that I will walk to work tomorrow (as I even now intend), but not that I will
trip over a child’s toy and break my leg (unintentionally). In the “will_{BF}”-sense, both
propositions are true. Now, in cases where one can move transparently from \textit{I will} \phi to \textit{I
intend to} \phi, is this “will_\textit{I}” or “will_{BF}? Certainly the step is warranted if the relevant “will” is
“will_\textit{I}”: in this case, the thought \textit{I intend to} \phi just unpacks part of what the subject is already
committed to in accepting that she will \phi. But the step looks much harder to understand if
her basis is sheerly a conviction that she will_{BF} \phi: in that case, she believes that her \phi-ing will
occur, but what sort of reason is this for believing that she now intends to \phi? Before I met
my wife, I believed that I would someday meet the woman who was right for me – and, I can
assure the reader, not on the basis of any induction from past experience or other
conviction that there was good evidence for this. It looks as though Byrne’s \textbf{INT}-schema
should have warranted me in inferring: I intend to meet the woman who is right for me. But
I was not so deluded as to suppose I could \textit{intend} any such thing, though of course I
intended to be open to the possibility of this happening. Surely there are plenty of things
we believe will occur in our lives, without supposing ourselves to have good evidence,
though we do not now intend them to occur. This suggests that Byrne’s \textbf{INT}-schema is not a
reliable principle of inference after all, and achieves its appearance of plausibility by trading
on an ambiguous use of “will”, one that allows us to slide uncritically toward the charitable
reading of the premise: I will_\textit{I} \phi.

If my normal basis for judging that I intend to \phi is a conviction that I will_\textit{I} \phi, then
although my basis is superficially neutral, it is not genuinely neutral: it already expresses a
(purported) awareness on my part of what I intend to do. My basis is a judgment about the
world, but not \textit{sheerly} about the world. I am, we might say, already thinking of the future as
“mine to settle” in making this sort of judgment about what will be. I think this is true more
generally in transparent self-knowledge: the thought about the world from which one
departs is not genuinely neutral. It is not a sheer proposition about the world, but a way of
representing the world that presupposes something about the mind of the representing
subject. I will make a case for this interpretation of transparent knowledge of our own
perceptions and beliefs, and will further elaborate the distinction between sheer

virtue of her present intention. We can usually tell from context whether this is meant.
propositions about the world and mind-presupposing propositions, when I turn to my own view below.

For the moment, let me note a final reason for skepticism about Byrne's INT-schema, when its premise is read as a proposition sheerly about the future of the world. This account of the relation between my awareness that I intend to φ and my conviction that I will φ seems to me palpably self-alienated. In certain cases, I believe that I will φ precisely because I knowingly intend to φ. It is on the basis of understanding my own power to choose to do things, and in light of my thinking about what to do, that I form these convictions about what I will do. These views of the future express “practical knowledge” (or at any rate, practical conviction): they do not involve a merely speculative or spectatorial view of my future, but a stance toward my future as mine to determine. The idea that the line of epistemic dependence runs in the other direction – from a sheer conviction about what I will do to a conclusion about what I now intend to make the case – leaves my knowledge of my own intention looking spectatorial in just the sense that Moran, following Wittgenstein, rightly repudiates. For such knowledge would be grounded, not in my seeing a certain act as in my power and regarding it as to be done, but in my supposing that this is how things will pan out. It seems to me that a subject who holds a view about what he intends on this sort of basis does not consciously intend. And again, I believe this is an instance of a more general problem with Byrne’s view. If our knowledge of our own minds were grounded in sheer propositions about the non-mental world, it would in general be merely spectatorial knowledge of our own mental states, not, as it were, knowledge of them from an engaged perspective. The resulting convictions might be true, but in voicing them I would not speak from the standpoint of the relevant mental states. We can put this by saying that I would not speak from a consciousness of those states.19 This is my third objection.

4. Peacocke on transparency and conscious judgment

A lesson we might draw from our discussion of Byrne is that the basis of (apparently) transparent self-knowledge really consists in some already-existing awareness of our own

---

19 I introduce “consciousness” here simply as a label for the relation to one’s own mental states that I have been trying to describe. It is not in general equivalent to “phenomenal consciousness” as this is standardly characterized, nor does it amount to mere “access consciousness”, on the standard use of this term. It does, I believe, have a basis in (some) ordinary usage of the term “conscious”, and it is a sense of “consciousness” relevant to Sartre’s use of the term, which we will discuss below.
minds. Christopher Peacocke first articulated his account of doxastic transparency before Byrne had written on the topic (Peacocke 1998), but his view can be thought of as contrasting with Byrne’s in just this way: he locates our ground for making self-ascriptions of belief in our capacity to make conscious judgments. Since it is crucial to Peacocke’s view that judgment is a kind of “phenomenally conscious” event, a mental happening whose occurrence contributes to the phenomenal character of the subject’s conscious state, and since he holds that the presence of this distinctive phenomenal character is crucial to her warrant for self-ascribing belief, his view is not that the subject’s basis for self-ascribing belief is a sheer (apparent) fact about the non-mental world. 20 In one respect, I think this is a step forward, but in another, I want to argue that Peacocke’s view is a step back. Seeing where the view goes awry will help us to see the specific merits of reflectivism.

Peacocke’s account of doxastic transparency is set in the context of a complex and systematic body of views, developed over decades, about the nature of epistemic warrant, concepts and concept-possession, reference, and so on. I cannot take up these views here. I believe, however, that we can extract from his work a relatively straightforward account of doxastic transparency, one that is less subtle than Peacocke’s full view, but that captures a major part of what makes his approach intuitively attractive.

The simplified Peacockean view (which I will henceforth just call “Peacocke’s view”) is this. As a rational subject, I can normally consider whether p and answer this question by judging that p (or that not-p, that it is unclear whether p, etc.) 21 Judgment is a phenomenally conscious event, hence one of whose occurrence I automatically have a kind of awareness. 22 So when I judge that p, I am aware, not merely of the (apparent) fact that p, but of my so judging (or at any rate, of a phenomenal profile characteristic of my so judging). This in turn warrants me in judging that I believe that p, since judging that p normally either expresses a belief that p (if one already exists) or else produces a belief that p (if one does not yet exist). These connections hold in the normal case, but they need not hold universally: sometimes an act of judgment does not derive from, or result in, a standing state of belief. So phenomenal consciousness of judging that p is a fallible indicator that one

---

20 Nico Silins has recently defended a similar view, which he explicitly contrasts with Byrne’s approach in this respect. Cf. Silins 2012, pp. 304, fn. 12 and 306, fn. 17.

21 Note that this is true whether or not my answering the question whether p requires “making up my mind”. Even if my judgment expresses a preexisting belief, I answer my question by judging.

22 Peacocke suggests that it is specifically an “action awareness” (Peacocke 1998, p. 88, elaborated in Peacocke 2008), but this will not be crucial for the issue that concerns us.
believes that $p$, but it is an indicator that is normally reliable. Hence, when the subject has no special reason to doubt that her situation is normal, and as a matter of fact the resulting self-ascription of belief is true, the subject who self-ascribes a belief on this basis thereby comes to know what she believes. 23 And though there may be cases in which a subject forms a belief about what she believes without an intervening act of conscious judgment, even in this kind of case, her warrant for doing so rests, not on sheerly her belief that $p$, but on the fact that she would have judged that $p$ if she had considered the question. 24

This seems to me a natural alternative to Byrne’s interpretation of doxastic transparency. It rejects the idea that the subject’s basis for self-ascribing a belief is that $p$, and holds instead that her basis is a conscious mental event: her judging that $p$. This is certainly a possible interpretation of the phenomenon Evans described: it might be that I learn what I believe by considering a question about the wider world, but this look outward warrants me in self-ascribing a belief only because it results in a judgment, my consciousness of which warrants me in drawing a conclusion about what I believe. This indeed is how Peacocke sees the matter: he says that “[t]he description of a self-ascription made on a particular occasion as consciously based should not be regarded as in competition with” the method described by Evans, since, just as Evans says, it is by “putting into operation my procedure for answering the question whether $p$” that I come to have a basis for self-ascribing a belief (Peacocke 1998, pp. 72-3).

In spite of its naturalness, I think this account mischaracterizes something crucial about our cognitive relation to our own beliefs. The essential structural features of the account are that (i) it requires a factor distinct from the (apparent) fact that $p$ as an element in the subject's warrant for self-ascribing belief, and (ii) this further factor is not itself an awareness of believing that $p$, but an awareness of something that is only a fallible indicator of belief. We can see what is odd about this proposal by once again considering the matter from the standpoint of a subject trying to understand her reason for taking it to be true that she believes that $p$. What can she say to herself? I take it there are two kinds of answers she might give, corresponding to two readings of Peacocke’s appeal to conscious judgment. On the one hand, she might appeal directly to the fact that she has judged that $p$. On the other hand, she might appeal to the fact that she has experienced a conscious state with a distinctive phenomenal character, the “special feel” of judging that $p$, as it were.

On the first option, the subject’s awareness that she believes \( p \) is supposed to be grounded in her awareness that she has judged that \( p \). It is therefore important that her awareness of judging that \( p \) should not itself require awareness of believing that \( p \): if it did, it could not be the basis of her awareness that she believes that \( p \), but would rather depend on this very awareness. So it is crucial to Peacocke’s view that the step from conscious judgment to knowledge of belief is a fallible step from one item of awareness to another distinct item of awareness. But I think we should find this conception of conscious judgment hard to understand. After all, not just any event of consciously entertaining the content that \( p \) is a case of judging that \( p \). If I entertain the idea of \( p \)'s being the case noncommittally (merely hypothetically, in a counterfactual spirit, etc.), I have not thereby expressed conviction in the truth of \( p \), and so presumably I have not judged. Judging that \( p \) requires not merely inwardly affirming that \( p \) (whatever that might mean) but affirming \( p \) in the conviction that it is true. But it is hard to see how this can mean anything less than: it requires inwardly expressing one’s belief that \( p \). And then it is hard to see how I can take myself to judge that \( p \) without presupposing that a certain conscious event expresses my taking \( p \) to be true. So my awareness of judging that \( p \) does not seem to provide me with an independently available ground for believing that I believe that \( p \). If I am fallible about whether I believe that \( p \), then for the same reason I am fallible about whether I (genuinely) judge that \( p \): my warrant for the latter determination must include my warrant, whatever it may be, for the former.

There is a kind of example that convinces some philosophers that it must be possible consciously to judge that \( p \) without believing that \( p \). Peacocke offers a widely-discussed instance of the genre:

Someone may judge that undergraduate degrees from countries other than their 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. (1998, p. 90)

I agree that such phenomena can occur, but I do not think they show that one can consciously judge that \( p \) while not believing that \( p \). If anything, they show that consciously judging that \( p \) is consistent with believing that not-\( p \), but it is not impossible for a subject both to believe that \( p \) and to believe that not-\( p \) (though such a conflicted state is certainly a cognitive pathology). Such examples may seem to show more than this because the subject’s judgment that \( p \) is described as not expressing a lasting conviction, and this may
seem to entail that it does not express a *state* of belief. But although it is true that belief is a state, not all states are lasting. A state is a mode of being that is *such as* to last, other things equal, but other things are not always equal. Perhaps the person in Peacocke’s example has a wavering belief that undergraduate degrees from other countries are of an equal standard, one that comes when she considers the reasons but dissipates thereafter. Or perhaps she never genuinely believes this at all, and does not genuinely judge it to be so. I do not deny that one can be mistaken about whether one really believes something: what I dispute is that this possibility speaks in favor of an epistemology of belief on which knowledge of belief is mediated by another, independently-available form of awareness.

At this point, it may seem appealing to shift to the other interpretation of the subject’s basis for self-ascribing the belief that *p*: not an awareness of judging that *p* as such, but of a state of consciousness with the characteristic phenomenology of judging that *p*. This, we might argue, can be present both when a person judges that *p* in the belief that *p* (or as I’d prefer to say: genuinely judges that *p*), and also when a subject judges that *p* without believing that *p* (or as I’d prefer to say: merely seems to herself to judge that *p*). So this can supply her with a fallible, non-question-begging basis for a self-ascription of belief.\(^{25}\)

Now, in entertaining this option, I am trying to play along with the use of concepts I do not really understand: I am not clear what it means to say there is a “phenomenology” of judgment, over and above saying that on some occasions when I take myself to express conviction in a certain proposition, I do not actually express such conviction. But let us suppose there is such a thing: what it is like for me (seemingly) to judge that *p*. Could this give me a reason to self-ascribe the belief that *p*? I do not understand the notion well enough to argue that it couldn’t, but it seems to me that if it did give me a reason, it would at best put me in a position to have the kind of knowledge of my own belief that I have called “spectatorial”. The relevant phenomenal profile is supposed to be one I can be aware of while leaving it open whether I genuinely believe the proposition in question: it is a normally reliable indicator of belief that *p*, but is consistent with the absence of this belief. On the basis of this indication, in the absence of a reason to believe that my situation is abnormal, I am supposed to be warranted in self-ascribing the belief that *p*. Do I believe *p* to be true? Well, very probably. That is a speculation about myself, not a conscious expression of conviction that *p*. What I can acquire by this method, if anything, is a bit of information

about myself, not a self-consciously held stance on the question whether \( p \). In locating my basis for self-ascribing belief in a fallible indicator of belief, Peacocke’s account drives a wedge between my consciously taking it to be the case that \( p \) (which I express in judging that \( p \)) and my knowledge that I believe that \( p \) (which I derive from, but am not entitled to identify with, my conscious assessment of \( p \) as the case).

If this is right, then Peacocke’s account has lost hold of the true spirit of Evans’s observation, even if it fits the letter of his account. It has lost hold of the idea that my transparent knowledge of my own beliefs is expressive of how the world looks to me. Such knowledge is not merely arrived at by considering whether \( p \); it remains a mode of knowing in which I (self-consciously) look outward. This, I believe, is the deeper sense in which our knowledge of our own minds can be transparent: not merely that it can be based on a consideration of the world, but that it can consist simply in a self-aware stance toward the world, not an independent knowledge of one’s holding such a stance.

5. The reflectivist approach to transparency
I have objected to the idea that a subject’s basis for transparent self-knowledge is a sheer proposition about the world, and also to the idea that her basis is some characteristic phenomenal state that serves as an indicator of her own belief. For different reasons, I have argued, each of these accounts could at best provide the subject with an alienated, spectatorial knowledge of her own mind, but this conflicts with the lesson we took from Moran: that our capacity for transparent self-knowledge is a marker precisely of our having an engaged, non-alienated awareness of our own mental lives. I conclude that neither of these accounts of transparency is satisfactory.

I believe, however, that there is an insight worth preserving in each approach. Our discussion of Byrne suggested (at least in the case of intention – but I think the point generalizes) that a satisfying account of transparency would need to allow that a subject makes transparent self-ascriptions of mental states on the basis of some sort of antecedently-existing awareness of her own mind. Our discussion of Peacocke suggested (at least in the case of belief – but I think the point generalizes) that the relevant awareness of the subject’s own mind must be, not a mere indicator that one is in the relevant mental state, but an “engaged” awareness of that mental state: one that does not introduce a distinction between knowing the relevant mental state and holding the world to be as that state represents it as being. A transparent awareness of one’s belief that \( p \) must be an
awareness *in which* one takes it to be the case that ϕ; a transparent awareness of one’s intention to ϕ must be an awareness *in which* one is determined to ϕ; and so on. So, to put the matter roughly, Byrne is right in his resolute insistence that transparent self-knowledge must look outward, while Peacocke is right to think that its basis must not be a sheer awareness of the world. But can there be a kind of awareness that satisfies both of these demands?

In earlier work (Boyle 2011b), I tried to argue that there can. The interpretation of transparent self-knowledge that I proposed, which I called “reflectivist”, holds that we are warranted in self-ascribing various sorts of mental states on the basis of a consideration of the world because there is already a kind of self-awareness implicit in the relevant ways of representing the world. I put this by saying that the existence of the relevant mental states (perception, belief, intention, etc.) itself involves the subject’s tacit knowledge of their existence, so that all the subject needs to do to achieve explicit knowledge of these states is to *reflect* on what she already knows. On this view, the subject’s step is not, as Byrne suggests, an inference from one item of knowledge to another distinct item of knowledge; nor is it, as Peacocke proposes, a transition from phenomenal consciousness of one mental event to propositional knowledge of another, distinct mental state. It is not the advent of new knowledge, but simply the subject’s articulation for herself of the significance of an awareness that was already involved in her regarding the world in a certain way. I tried to argue that, on a charitable reading, Moran can be understood as offering an account of the nature of belief that satisfies the key demand of reflectivism: that belief itself should involve the subject’s tacit knowledge of believing. And I suggested that we might hope to give accounts of other kinds of mental states that, while different in their details, also met this demand.

I continue to think that only an approach along these lines can accommodate both Byrne’s insight and Peacocke’s insight, but I am no longer satisfied with the account I gave of reflectivism. Whether the reflectivist approach is acceptable will depend in large part on whether plausible stories can be told about how various specific kinds of mental states meet the reflectivist demand. But there is also a kind of worry about the approach that arises even before we turn to specifics: a concern about the clarity, and perhaps the coherency, of the notion of “tacit knowledge”. The very term may seem to withdraw with the adjective what it grants with the noun. For if believing (to focus for simplicity on a specific state) involves *knowledge of* believing, must that not mean that it involves the subject’s
representing it as the case that he believes that \( p \)? But then in what sense can his knowledge of his own belief be tacit, other than the inconsequential sense that he has not yet verbalized this representational content? There are contemporary philosophers – most famously, Sydney Shoemaker – who hold that, in the right conditions, believing that \( p \) simply constitutes knowing that one believes that \( p \).\(^{26}\) But on this “constitutivist” view, such a belief is simply known, not “tacitly” known. The reflectivist view, by contrast, maintains that, before a subject reflects, there is a sense in which her knowledge of her own belief is not yet present as such, but merely latent in her manner of representing the world. Indeed, this is crucial to the reflectivist’s attempt to give the phenomenon of transparency its due: he seeks – I seek – not simply to maintain that when one is in a world-directed mental state, one also knows that one is in that state, but to respect the thought that our reflective knowledge of our own mental states is in some way grounded in our world-directed awareness.\(^{27}\) But is there room in logical space for such a position? If a subject who believes already knows that she believes, in what sense can she not yet know this “explicitly”? Or if she does not yet know it “explicitly”, in what sense can she already know it, and how can this knowledge ground her reflective self-ascription of belief?

The notion of tacit knowledge is also crucial to the reflectivist approach for another reason: it allows the reflectivist to avoid imposing implausibly strong intellectual requirements on belief (and also on perception, intention, etc.). For explicit knowledge is commonly assumed to be subject to the following Concept Possession Requirement:

\[(\text{CPR}) \quad \text{A subject can know that } p \text{ only if he possesses the concepts necessary for understanding the proposition that } p.\]

If the reflectivist claimed that belief requires explicit knowledge of belief, where explicit knowledge is subject to (CPR), he would be committed to holding that a subject can have beliefs only if she possesses the concept of belief. But surely this would be an excessively strong requirement on belief.\(^{28}\) The concept of belief seems to be a special and sophisticated attainment, a first step toward a philosophical account of belief, whereas it

---

\(^{26}\) For his most recent statement of his view, see Shoemaker 2012.

\(^{27}\) In conversation, Alex Byrne has queried whether a view like mine can respect this. I’m grateful to him for forcing me to see the difficulty.

\(^{28}\) Shoemaker (2012) avoids the implausibility in another way: by holding that belief constitutes knowledge of belief only in a subject who possesses the first person concept and the concept of belief, whose belief is available for use in explicit theoretical and practical reasoning, etc. But the reflectivist aims to make a claim about the nature of belief as such (at least the sort of belief that occurs in rational creatures, ones capable of rational deliberation about what is the case).
seems that quite unreflective persons, not to mention infants and nonhuman animals, can hold beliefs and act intelligently on the basis of them. So the reflectivist had better hold that the knowledge of belief implied in belief is not explicit knowledge. But if we sever the link between ascribing knowledge that one believes that \( p \) and ascribing grasp of the concept belief, we make it much less clear what this ascription comes to. Just what is at stake in saying that the subject "tacitly knows" this?

6. Sartre and reflectivism

Having raised these difficulties for myself, let me note that I am not alone in holding a position with this structure. Sartre also held such a view, though his terminology differs from mine. Whether this is good company is debatable, but at any rate, my idea was inspired by his, and I have come to think that I should have followed him even more closely.

The Sartrean doctrine that inspired me was his distinction between what he calls “positional” and “non-positional” consciousness. Sartre famously accepts the following propositions about consciousness:

\( (S1) \) “All consciousness [conscience] ... is consciousness of something. This means that there is no consciousness that is not a positing [position] of a transcendent object” (BN li/17).

\( (S2) \) “[All knowing consciousness [conscience connaissante] can be knowledge [connaissance] of nothing but its object” (BN lii/18).

\( (S3) \) “[All positional consciousness of an object is at the same time non-positional consciousness of itself” (BN liii/19).

\( (S4) \) “It is the non-reflective consciousness [conscience non-réflexive] that makes reflection possible” (BN liii/19).

(S1) is a point that Sartre credits to Husserl: it encapsulates the idea that the defining trait of the psychic aspect of our existence – "consciousness" being Sartre's generic term for the mode of being of the psychic, a term not restricted to occurrent states – is its intentionality, its being of or about some object that is not identical to the consciousness itself. In this sense, consciousness “transcends” itself to posit a realm of being beyond itself. “Positing” is...
Sartre’s term for the relation of consciousness to its objects: consciousness is “positional” to the extent that it is of or about an object. Now, Sartre does not think that all positional consciousness consists in knowing an object – there are other modes of positing, such as imagining, desiring, and so on – but Sartre does regard knowing as a species of positional consciousness: it is a kind of relation in which consciousness stands to some posited object. That object, and only that object, is what the relevant consciousness is knowledge of. Thus (S2).

The crucial point for my purposes is (S3): all positional consciousness of an object involves non-positional consciousness of that state of consciousness itself. “Non-positional” consciousness is a mode of awareness that does not posit that of which it is aware as its intentional object. This may sound like a paradox: how can there be an awareness of something that does not posit that thing, if to posit a thing just is to relate to it in such a way that one’s consciousness is of it? Sartre argues, however, that there must be such non-positional consciousness of consciousness if there is to be positional consciousness of objects:

[T]he necessary and sufficient condition for a knowing consciousness to be knowledge of its object is that it be consciousness of itself as being this knowing. This is a necessary condition: if my consciousness were not consciousness of being conscious of the table, then it would be consciousness of the table without consciousness of being such, i.e., a consciousness ignorant of itself, an unconscious consciousness—which is absurd. It is a sufficient condition: that I am conscious of being conscious of this table suffices for me in fact to be conscious of it. (BN lii/18)

Sartre thus thinks the idea of a consciousness that is not (non-positonally) conscious of itself is absurd: this would be an “unconscious consciousness”, which is a contradiction in terms.

I suspect most contemporary philosophers would not think much of this argument: they would reply that it is one thing to be conscious of an object, another to be conscious of one’s own state of consciousness, so the supposed contradiction here is specious. But we

---

30 I will follow the common practice of speaking of “states” of consciousness, though Sartre himself would reject this mode of expression as implying a kind of passivity that is foreign to consciousness (cf. The Transcendence of the Ego, pp. 61-8/45-51, 109n/15n). I cannot explore this issue here, but it will be useful to have some common noun designating the sort of thing exemplified when a subject is conscious of something, and I think the term “state” is innocuous once its potentially misleading connotations have been flagged.

31 Sartre is aware of the difficulty, and responds by placing the “of” in parentheses when he speaks of non-positional consciousness (of) something (cf. BN liv/20). But it is not clear that mere parentheses can ward off a contradiction.
will see shortly that it is possible to make a more forceful case for (S3) than Sartre does here. For now, the point to keep in mind is that, for Sartre, knowledge of an object depends on a kind of self-consciousness that is different in structure from knowledge of an object. This point applies even in the case of our \textit{reflective} knowledge of our own states of mind: when we posit our own mental states as objects of knowing consciousness, but this reflective knowledge depends on a more basic non-positional consciousness of these states. Thus (S4).

My “reflectivist” approach was modeled on Sartre’s view in that I also held that our capacity for explicit, reflective knowledge of our own mental states is grounded in another, more basic sort of awareness of these states, one that is involved in their very obtaining. I called this more basic awareness “tacit knowledge”, aiming by the inclusion of “tacit” to capture the Sartrean idea that this awareness is not positional, but is an immersed mode of awareness that leaves the focus of cognition, not on the mind, but rather on the world. But I have come to think it better to follow Sartre in calling this awareness “consciousness”, reserving the term “knowledge” for an awareness that makes that of which it is aware its intentional object and brings some concept to bear on it. To suggest that (e.g.) belief involves \textit{knowledge} of belief courts all the difficulties noted earlier, and adding that the relevant knowledge is “tacit” does not make clear how to avoid them. Using the term “consciousness” does not itself provide the needed clarification, but it at least marks the spot where clarification is required.

Sartre’s idea is that there is a kind of consciousness whose intentional object is an aspect of the non-mental world, but which can have its specific world-directed significance only in virtue of presupposing a non-positional consciousness of that very state of consciousness. I want to suggest that it is just this sort of non-positional consciousness whose presence makes transparent self-knowledge possible. We can see the need to recognize such consciousness by seeing how various ways of representing the world depend on it, and how our transparent knowledge of our own mental states draws on it.

As a first illustration, consider transparent knowledge of one’s own intentions, which we discussed earlier in connection with Byrne. I granted Byrne’s idea that we can be warranted in ascribing an intention to \(\phi\) to ourselves simply in virtue of accepting a certain sort of proposition about the future:

\begin{equation}
(I1) \quad \text{I will } \phi.
\end{equation}

But I argued that if this inference is to be reliable, and to give me non-alienated knowledge
of my own intention, the “will” in (I1) must be, not “willBF”, but willi: a way of thinking of
the future that involves a non-theoretical stance toward it, as it were. It is a way of thinking
of the future as mine to determine, indeed as settled by my so thinking – as comes out in the
kinds of grounds on which we accept such propositions about what we will do.
Nevertheless, it is a way of thinking of the future of the world. There is no explicit content
pertaining to my present mental state. Nor can my thought be factored into a proposition
about my present mental state and a sheer proposition about my future, as in

(I2) I intend to φ & I willBF φ

(I2) is naturally read as expressing an alienated, predictive attitude toward my future φ-ing:
I take it that I now intend it, and, as a distinct matter, I accept that it will occur. This is not
equivalent to (I1), when (I1) is read as I willi φ, for (I1) represents this aspect of the future
as hereby determined by me. We have in (I1) a distinctive way of thinking about the future,
not merely a generic way of thinking about the future plus an implicit claim about my
present mental state. We have, to put the matter in Sartrean language, a mode of
consciousness of the world that is possible only in virtue of non-positional consciousness of
one’s own consciousness. And although this implicit consciousness does not itself involve
an application of the concept of intention, it warrants a self-ascription of intention: the
application of this concept just makes reflectively explicit a consciousness of my hereby
determining to φ that was implicit in my way of thinking of the world.

Turn now to transparent knowledge of one’s own perceptions. Consider a subject
who makes a transition from the world-oriented observation

(P1) This cat is purring.

to the reflective thought

(P2) I perceive a purring cat.

How can this be a reasonable transition to make, given that there might very well be a
purring cat here though I do not perceive it? Well, consider the way the cat is represented
in (P1): it is represented as available for demonstrative reference, expressible by using a
“this” and a common noun to form a singular term. Now, an object is available to a given
subject for demonstrative reference only in specific kinds of circumstances. Though the cat
may be purring, I cannot successfully think of it as this cat when it is miles away, or hidden
behind a screen, or known to me only by hearsay. Demonstrative expressions are used for
other purposes, but there is a recognizable and distinctive use of them on which they mark

32 This needs more argument, but I hope at least to have conveyed a strategy for making such a case.
a mode of availability for thought grounded precisely in the fact that the subject perceives the object in question, and this is the natural reading of the "this" in (P1).

Now, a subject who thinks the thought expressed by this reading of (P1) is not thinking about her own perceptual state: the object of her thought, if it is well-grounded, is a certain cat in her vicinity. But the aptness of her way of thinking of the cat will depend on her perceiving it, and she will have the capacity to think of things in this way only inasmuch as she is in general able to recognize the kinds of circumstances that make such thoughts apt. Moreover, her way of thinking of the cat is not reducible to a proposition about her own mental state plus a sheer proposition about her environment, as in

(P3) A purring cat is here and I perceive it.

(P3) does not capture the singular mode of presentation of a cat expressed in the "this" of (P1). It represents the subject as having a nonsingular, merely existential mode of awareness of a cat, thereby losing the way in which perception makes possible a distinctive mode of awareness of the cat. So again, we have a mode of consciousness of the world that is possible only in virtue of non-positional consciousness of one's own consciousness. If the subject goes on to think the reflective thought

(P2) I perceive a purring cat

she will be making explicit a psychological state whose presence was already presupposed in her world-directed representation of the cat. What justifies her reflective step is not the sheer fact that a certain cat is purring – supposing that is so – but her non-positional consciousness of her own manner of apprehending this fact.

Consider finally the case of belief. This seems to me the most difficult case, and I am least confident of what I have to say about it, but I think a version of the Sartrean point holds here too. Suppose I am considering the question whether there will be a third world

---

33 Might she merely be sensitive to circumstances that are in fact ones in which she is perceiving a certain object, but not recognize those circumstances as such? Well, I have granted that her thinking of an object as this F does not involve her representing herself as perceiving the relevant F. Her mode of representation does, however, presuppose that she stands in a specific kind of relation to relevant F (viz., a perceptual one), and her sensitivity to just this relation will come out in the specific kinds of doubts she is prepared to entertain about such representations, the ways she checks their soundness, etc. In these ways, she shows an awareness, implicit but open to reflective articulation, of the specific kind of relation in which she stands to the object of her representation.

34 The presupposition of her thought may of course be false: her representation this cat may express a merely apparent awareness of a perceptually-presented cat. Much more would need to be said in a full account of our warrant for self-ascriptions of factive and non-factive perceptual states. More would also need to be said to account for the ways in which we know the sensory modality of a given perception, etc. I am just trying to illustrate the basic strategy here.
war, and I reach the conclusion that

(B1) There will be a third world war.

What I conclude is a proposition about the future of the non-mental world. But this proposition stands in specific relations to various other propositions I might have affirmed. For instance, it presupposes

(B2) It is possible that there will be a third world war.

And it excludes

(B3) There will not be a third world war.

(B1) also contrasts with other, less committal propositions about the purport of the evidence on which I base my judgment. I might, for instance, have concluded

(B4) The available evidence suggests that there will be a third world war.

while leaving it open that this evidence may not be conclusive.

Now, a subject who can deliberate about factual questions will be a subject who can consider the propositional question whether \( p \) as open and can entertain a range of possible answers to it, who can assess the strength of her evidence for a given answer, and so on. Her concluding that (B1) will express her taking this to be the correct answer to the question whether there will be a third world war, in the understanding that contrasting answers such as (B2)-(B4) are possible. Moreover, it seems to me that her concluding that (B1) must involve an implicit awareness of her taking this answer to be correct, for if she were not aware of this – if her representing (B1) did not express a consciousness of having resolved the question whether there is a third world war – then the question would still remain open for her, and her deliberation would not have concluded. So although what she represents as the case is a proposition about the non-mental world, her manner of representing it depends on an implicit awareness of her own determination about what is correct. We might therefore say that in concluding that there will be a third world war, she expresses a non-positional consciousness of her own conclusion: an awareness that figures, not as the object of her thought, but as the necessary background of her thinking of the question of whether there will be a third world war as settled. If she then goes on to think the reflective thought

(B5) I believe that there will be a third world war

she will be bringing to bear a psychological concept in a way that makes her attitude on the question explicit, but an awareness of this attitude was already implicit in her world-directed representation of a third world war, not merely as something that might take place
or that looks likely to take place, but that will take place. What justifies her reflective step is not the sheer fact that there will be a third world war – supposing that is so – but her non-positional consciousness of her own stance on this question.  

7. Conclusion

Let me conclude by noting some advantages of this approach and some topics for future work. I have already argued that the reflectivist approach allows us to accommodate both Byrne’s insight that transparent self-knowledge is grounded in a consideration of the world and Peacocke’s insight that it must rest, not on a sheer awareness of the world, but an awareness already informed by some sort of consciousness of my own mind. It allows us to explain how a look outward can itself make rational a judgment about one’s own mental state. By the same token, the reflectivist approach explains why Moran is right that transparent self-knowledge is a marker of a non-spectatorial awareness of our own mental states. For according to reflectivism, transparent self-knowledge just makes explicit a mode of awareness that is already involved in our immersed stance toward the world. It is not a knowledge of my own mind that holds open the question whether the world is as my mental state represents it to be, but simply a self-conscious look outward.

An advantage of drawing the Sartrean distinction between immersed, non-positional awareness of our own minds and reflective knowledge of our mental states is that it puts us in a position to respond to the concern that reflectivism imposes implausibly strong intellectual requirements on belief, perception, intention, etc. We can admit that a subject might (e.g.) believe that p without possessing the concept of belief, for on our Sartre-inspired view, the awareness of believing involved in belief does not take the form of a positional representation of oneself as believing, but of a non-positional awareness implicit in a certain manner of representing a proposition as so. We can also take a concessive attitude toward cases like Peacocke’s administrator who believes that undergraduate degrees from other countries are not up to the standard of his own country, though he does not recognize himself to believe this. Since it is one thing to have non-positional awareness of believing and another reflectively to think that one believes, it is possible in principle for

---

35 Note that, though we have developed this point with reference to a case in which a subject reasons to a conclusion, an appeal to reasoning was not essential to our account. What was important was that the subject holds the views he does in a way that involves non-positional awareness of himself as taking a stand on a question on which various stands are possible. In a subject capable of deliberation, this is true quite generally, whether or not he holds these views on the basis of reasoning, indeed whether or not he holds them for specific reasons at all.
a subject to believe and yet not think himself to believe, or indeed while thinking it wrong to believe as he does.36

We are in a position to make these replies because we have granted that the basic form of self-awareness does not involve the application of a psychological concept to ourselves. The self-awareness included in belief does not posit the believer as an object or represent her condition as being of a certain kind – though it may put us in a position to frame on reflection the concepts of a belief and a believer. Once we recognize this point, we see the need for a new kind of inquiry into self-knowledge: not an investigation of how we are in a position to know our own minds at all, but an inquiry into the nature of the reflective act by which we transform a necessary non-positional awareness into an explicit knowledge of ourselves.

Sartre's Being and Nothingness is in effect an extended investigation of this topic, and it is worth noting how the distinction between non-positional self-awareness and reflective self-knowledge is connected with another great Sartrean theme, the perpetual threat of self-alienation that characterizes our lives. A major source of resistance to approaches that represent our mentality as essentially self-aware, I think, is the sense that they cannot do justice to the depth and ubiquity of this threat. But precisely because it distinguishes between non-positional self-awareness and reflective self-knowledge, Sartre's approach is built to recognize ways in which our reflective self-understanding can distort the reality of our psychic lives, both in the more prosaic sense that it can fail to appreciate the facts, and also in the more profound sense that it can give rise to a kind of "bad faith" in which there is no stable fact to know. An adequate treatment of this idea would need to consider how our capacity for reflection is the ground both of the temptation to bad faith, and also of the possibility of a project of authenticity. But these are topics for another day.

36 More would need to be said here. For now I just state the basic principle of the response.
References


