CANADIAN JOURNAL OF PHILOSOPHY
2019, VOL. 49, NO. 7, 1012–1039
https://doi.org/10.1080/00455091.2019.1565621

Routledge
Taylor & Francis Group

ARTICLE

Transparency and reflection

Matthew Boyle

Department of Philosophy, University of Chicago, Chicago, IL, USA

ABSTRACT
Much recent work on self-knowledge has been inspired by the idea that the 'transparency' of questions about our own mental states to questions about the non-mental world holds the key to understanding how privileged self-knowledge is possible. I critically discuss some prominent recent accounts of such transparency, and argue for a Sartrean interpretation of the phenomenon, on which this knowledge is explained by our capacity to transform an implicit or 'non-positional ' self-awareness into reflective, 'positional' self-knowledge.

ARTICLE HISTORY Received 1 November 2018; Accepted 30 December 2018

KEYWORDS Transparency; self-knowledge; Jean-Paul Sartre; reflection; privileged access; belief; intention

> '[T]he mode of existence of consciousness is to be consciousness of itself [*conscience de soi*]… [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*[1]

1. Transparency as phenomenon and as problem

In an influential discussion of first-person belief ascriptions, Gareth Evans observed 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?" (Evans 1982, 225)

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

CONTACT Matthew Boyle mbboyle@uchicago.edu Department of Philosophy, University of Chicago

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

> [A] subject can gain knowledge of his [perceptual] 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*… The result will necessarily be closely correlated with the content of the [perceptual] state which he is in at that time. (Ibid., 227–228)

The phenomenon Evans describes here has come to be known as the 'transparency' of (certain kinds of) self-knowledge. Our knowledge of our own mental states is said to be *transparent* inasmuch as we can knowledgeably answer questions about these states by attending in the right way, not to anything 'inner' or psychological, but to aspects of the non-mental world.[2]

Although Evans discusses only our knowledge of our own beliefs and perceptual appearances, there is reason to think his point may have 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 an important phenomenon here for which a theory of self-knowledge must account. The phenomenon is this: I seem to be able to know various facts about my own mind simply by considering questions about the non-mental world. To obtain this knowledge, I look, as it were, not inward but outward. I will refer to knowledge of my own mental states that is available in this sort of way as *transparent self-knowledge*.

Evans regarded his observations as demystifying our capacity for certain kinds of privileged self-knowledge. It can seem mysterious how we are able to say what we believe without observing ourselves, even though we must observe another person to determine what she believes; but, Evans 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.

Many subsequent writers, however, have thought of Evans's observations, not primarily as providing the solution to a puzzle, but as presenting

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

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 many values of *p*: I am very far from omniscient (or omnicredent, for that matter). And similarly, there are often states of affairs that hold in my environment although I do not perceive them: I am very far from omnipercipient. So what justifies me in answering such questions about my own psychology by looking to seemingly independent facts about the world?[5]

It will not suffice, as an answer to this question, 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-directed questions. This is certainly true, and it is part of what gives Evans's proposals their initial appeal, but it does not address the heart of the puzzle. 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 in front of me, I become cognizant of an apparent fact that is in fact the *topic* 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: 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.[6] It is this psychological knowledge whose warrant is in question, and the *problem of transparency* is that nothing in my apparent basis seems to supply a ground for it.

This problem of transparency has, in the last two decades, become a principal focus in discussions of self-knowledge, succeeding and to some extent supplanting more generic debates about the basis of 'first person authority', whether 'privileged access' is compatible with externalism about mental content, and so on. I believe this is a valuable development, not because those earlier debates were misconceived, but because focusing on 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 (viz., facts about our own present mental states), but how this variety of knowledge is related to our capacity to engage, theoretically and practically, with the non-mental world.

My aim in this essay is to propose a solution to the problem of transparency that speaks to this wider issue. The solution is inspired by Sartre's idea that all consciousness involves a form of 'non-positional' consciousness of our own consciousness. Sartre claimed, in effect, that our capacity to know our own minds is linked to our capacity to know the world because our awareness of the world constitutively involves a kind of implicit self-awareness, and that we draw on this awareness when we reflect. I want to suggest that this idea is the key to resolving the problem of transparency,

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

and that appreciating its relevance to this problem is the key to seeing the truth in Sartre's somewhat darkly-expressed point.

I came to see the importance of Sartre's notion of non-positional self-consciousness while trying to clarify a distinction I had been led to draw, in earlier work (Boyle 2011), between 'tacit' knowledge of one's own mental states and explicit, reflective knowledge of those states. In order to bring out the motivation for Sartre's position, it will help to explain what led me to draw this distinction, and why I now think it must be clarified along Sartrean lines. So I will begin by describing how I arrived at my earlier position, and will situate it relative to three other prominent approaches to the problem of transparency. Seeing the strengths and weaknesses of these approaches will help us to see the point of Sartre's idea.

2. Transparency versus alienation: Moran

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 compelling observations about the character of transparent self-knowledge, and because the other approaches we will consider are framed in 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*. It is not the case, he observes, that one can always determine whether one believes *p* by answering the question whether *p*: there are occasions on which we can only discover our actual beliefs by observing ourselves, much as a spectator would. These are, however, pathological situations in which we are 'alienated' 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 fact 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

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

> 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, 405)

Moran summarizes his proposal 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' or 'speculative' 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, 58, 63). In a slogan, our capacity to have transparent knowledge of our own beliefs rests on our capacity to 'make up our minds'.

It is hard to read Moran's work on doxastic transparency without feeling that he has put his finger on something important, 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 can be brought about through deliberation.[7] As we noted earlier, 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 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. Whatever explains our entitlement, in these cases, to treat our answer to the relevant psychological question as transparent to a corresponding world-directed question, it surely isn't that we constitute such states by deliberating.

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 of which we have transparent self-knowledge without (it seems) taking a deliberative stance toward them.[8] I believe, for instance, that the former U.S. President William Taft was born in my home state of Ohio, and doubtless I once had some basis for believing this, but I cannot now recall what it was. Did I see it on a plaque, or read it 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 believe without having specific grounds. Nevertheless, my knowledge that I believe this surely meets the Transparency Condition: I can answer the psychological question whether I believe Taft was born in Ohio simply by deferring it to the factual question whether Taft was born in Ohio. Yet 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

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

be true that I would 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, there is a crucial lacuna in Moran's proposal. Consider a case in which I deliberate about whether *p* and conclude: Yes, *p*. In so doing, I may in fact make it the case that I believe that *p*, but even if I do, what puts me in a position to *know* that I believe *p*? Moran's suggestion seems to be that my warrant for accepting *I believe that p* is connected with the fact that I have effectively concluded that *p is to be believed*, but it is not clear how this is supposed to carry me across the threshold to doxastic self-knowledge. Suppose for the sake of argument that my concluding 'Yes, *p*' is tantamount to my concluding: *p* is to be believed. This can surely be the case although I don't believe *p*. Indeed, as a imperfect being, this is presumably often my situation: there is something it would be right for me to believe, and yet I don't believe it. So it seems that the problem 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'? Moran's explanation of transparency thus does not close the crucial gap.

For all these reasons, Moran's account seems not fully satisfactory as it stands. Nevertheless, I believe it has at its foundation a compelling observation of which we should not lose sight. As we have seen, Moran contrasts the relation in which we stand to our own mental states when we can know them transparently with an alienated, 'spectatorial' relation to those states. He puts this point – focusing, as usual, on the case of belief – by observing that we can have transparent self-knowledge only of our conscious beliefs, and

> to call something a conscious belief says something about the *character* of the belief in question. It is not simply to say that the person stands in some relation of awareness to this belief… I see myself in this belief; my conscious belief forms the basis for my further train of thought about the thing in question. (Moran 1999, 188)

Moran goes on to argue that it is possible to know immediately and without observation that one holds a certain belief, and yet not *consciously* to hold the relevant belief. As an example, he asks us to consider an analysand who has so perfectly internalized the perspective of his psychoanalyst that he knows immediately what beliefs his analyst would attribute to him, and (rightly) ascribes those beliefs to himself, but does not consciously hold the beliefs in question.

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

I think there is a good point here, though the example is unnecessarily contentious. Whatever we think of the possibility of psychoanalytic knowledge of unconscious beliefs, we should admit that consciously believing *p* requires more than merely knowing, without observation or inference, that one believes *p*. To know that one believes (e.g.) that there will be a third world war requires only that one takes *one's own belief-state* to one to be of a certain kind. A person who *consciously* believes that there will be a third world war, by contrast, does not merely know herself to hold this view; she consciously inhabits the relevant viewpoint, in the sense that, when she thinks about the world-oriented question whether there will be a third world war, it seems to her that there will. Moran's insight is that transparent self-knowledge of a mental state is available just when the relevant mental state is conscious in this sense: the point of view one ascribes to oneself is the very one that one consciously inhabits, in such a way that the question of what one believes and the question of what is so fuse into one.

It seems to me that this point embodies an insight that is independent of Moran's more specific claims about the explanation of doxastic transparency. The insight is that transparent self-knowledge is not merely knowledge that can be had immediately and without self-observation; it is knowledge grounded in consciously inhabiting the relevant point of view, and as such contrasts with the knowledge of a spectator. A spectator might know that I believe that *p*, but it is one thing for her to know this about me and another for *her* consciously to see the world from the perspective of *p*-believer. If unconscious belief is possible, then it is also possible for me to stand in this sort of spectatorial relation to one of my own beliefs: I can know myself to believe that *p*, and yet not consciously inhabit the perspective of a *p*-believer. But at any rate, for a person who consciously believes that *p*, being aware of her own belief and seeing the world from the perspective of *p*-believer are two aspects of the same awareness. Such a person does not merely have a point of view on the world, on the one hand, and know of its existence, on the other. She consciously holds the relevant point of view, and can thus express both *belief that p* and *knowledge that she believes that p* in a single act, by saying 'I believe that *p*'. We can thus say that she knows her own belief from the perspective, not of a spectator, but of an inhabitant of the relevant point of view

This point can be accepted by philosophers who reject Moran's idea that the only alternative to a 'spectatorial stance' toward oneself is a 'deliberative stance'. It may be that Moran overestimates the closeness of the tie between being able to know one's mental states 'transparently' and treating those states as open to deliberation. Perhaps this opposition characterizes only certain kinds of attitudes, if it applies anywhere. Be that as it may, it remains true in general

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

that our relation to mental states we can know transparently is not spectatorial. Perception can again serve as an example. People with normal perceptual awareness can, as Evans observed, have transparent knowledge of their own perceptual appearances. But consider a person with the kind of cerebral damage that produces 'blindsight'.[9] Blindsighted subjects do not consciously see what is presented in the 'blind' region of their visual field, but when prompted, they are able to perform perception-dependent tasks (for instance, making guesses about the features of objects in the blind field) with a better-than-chance rate of success. Such a subject stands in an alienated, spectatorial relation to her own seeing: she can say what she perceives only by drawing inferences from her own behavior (including her dispositions to make guesses about what is present in a certain region). But our normal relation to our visual appearances is not like this: it is not merely that we can know how things visually appear to us with an immediacy that is unavailable to the blindsighted person; it is that we consciously experience *the world* as containing objects with specific features, whereas she does not. Our knowledge of how things perceptually appear to us is thus not spectatorial: we are aware of these states *by inhabiting their perspective*, inasmuch as we consciously perceive the world around us to be a certain way.

At the same time, our knowledge of how things perceptually appear to us is obviously not deliberative: if we could deliberate about whether *to* have specific perceptual appearances, perception would not be the kind of cognitive power that it is. Moran's insight – that the availability of transparent self-knowledge marks a distinctively non-spectatorial mode of awareness of one's own mental states – thus applies more broadly than his emphasis on opposition between spectator and deliberator might lead one to suppose. This insight sets a condition of adequacy on accounts of transparent self-knowledge. I now want to argue that this is a condition some influential accounts of transparency fail to meet.

3. Transparency as inference from world to mind: Byrne

We observed earlier that there is a crucial gap in Moran's account of our warrant for moving from the judgment *p* to the judgment *I believe that p*: how does the former justify the latter? In seeking to fill this gap, there are two natural routes to explore. On the one hand, we might argue that our warrant derives somehow from what we consider: some (seeming) feature of the non-mental world. On the other hand, we might propose that it derives, not from what we consider, but from our consideration itself: some conscious mental state or process that occurs when we take up the relevant world-directed question. Each of these routes has in fact been explored, and I want to examine a prominent example of each approach – not simply for the sake of surveying

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

the literature, but because I think that each approach has significant attractions, but that each also misses something crucial.

The first approach I will consider seeks resolutely to defend the idea that, when one has transparent self-knowledge, the basis on which one ascribes a mental state to oneself is, not any kind of awareness of one's own mind, but a purported awareness of a fact about the non-mental world. As we have noted, this idea is not easy to accept, for it is hard to see how a such a step could be warranted. Nevertheless, if such an approach could be defended, it would provide an elegant solution to the problem of transparency, for it would explain our capacity for non-observational self-knowledge without appealing to any special faculty of introspection or other pre-existing form of self-awareness. It would thus clarify Evans's attractive idea that the relevant knowledge can be acquired simply by looking 'outward', rather than 'inward' toward our own mental states.

An approach of this sort has been forcefully defended by Alex Byrne (2005, 2011, 2018). Byrne's account rests on a simple idea: that we can acquire transparent self-knowledge by making an 'inference from world to mind' (Byrne 2011, 203). In general, Byrne holds, our capacity for inference is a capacity to make rule-governed transitions between acceptance of some set of propositions and acceptance of some further proposition. The problem of transparency is that the relevant inferences – for instance, the *doxastic schema*

BEL: *p*

I believe that *p*

look on their face as though they cannot yield knowledge, since their premises neither entail nor evidentially support their conclusions. But, Byrne suggests, this sort of objection to an inference-schema can be overcome if we can show that (1) inferring in accordance with the relevant schema reliably produces true beliefs and (2) such inferences are 'safe' in the technical sense that drawing this conclusion would not have produced false belief in any nearby possible world. If we could show this, and if we had no reason to suppose that BEL is *not* knowledge-conducive, we would, Byrne argues, be entitled to show deference to the view presupposed in our ordinary practice: that making such transitions is a way of coming to know what one believes.[10]

Byrne's project is thus first to identify inference-schemata corresponding to the kinds of mental states we can know transparently, and then to show that these schemata are 'neutral' (i.e., that their premises do not presuppose knowledge of the relevant mental states), reliable, and safe. If this can be shown, he holds, the problem of transparency will be solved, for we will have explained why the relevant inferences are, in spite of appearances,

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

normally knowledge-conducive. Moreover, we will in the process have accounted for the fact that we can speak with special authority about the relevant mental states, and we will have given an attractively economical account of this authority: one that does not appeal to any special introspective faculty, but only to general cognitive capacities required also for other kinds of cognition.

In the case of belief, the relevant inference schema is plainly BEL, and it is easy to show that this schema is safe and reliable. A subject will be in a position to infer according to BEL only if she accepts the premise: *p*. But this amounts to saying that she will be in a position to apply the rule only if she believes that *p*, and this ensures that her conclusion will be sound, whether her premise is true or false. In other cases the relevant inference-schema is less obvious, and the argument to its reliability and safety is less direct; but suffice it to say that Byrne makes a forceful case for the claim that, when their application is appropriately restricted, inferences such as the following are neutral, reliable, and safe:

SEE: $[\ldots x \ldots]_V$ & x is an F / ——— / I see an F[11]

INT: I will ϕ / ——— / I intend to ϕ[12]

By demonstrating that such analyses can be given in a wide variety of cases, Byrne (2018) seeks to show that the 'inference from world to mind' approach applies wherever transparent self-knowledge is possible.

Although I believe there is a great deal to learn from Byrne's analyses of particular relations of transparency, I think his general account of such transparency implies an unacceptably alienated, spectatorial picture of the subject's relation to her own mental states. To bring this out, I will raise two objections, which highlight different but related aspects this problem.[13]

First, it seems to me that Byrne's approach does not explain the rational intelligibility of these world-to-mind inferences from the subject's own standpoint. Our capacity for (personal level) inference is a capacity to arrive at new beliefs in virtue of seeing one or more (seeming) fact as supplying some sort of *reason* to accept some further proposition as true. The conclusions we reach through inference are not just beliefs that appear unaccountably in our minds; they are convictions for which we take ourselves to have a reason, and this is what makes them sustainable in the face of the capacity for critical scrutiny that belongs to rational subjects as such. What makes Byrne's BEL-inference problematic, it seems to me, is not merely the concern that it would be unreliable or unsafe, but a concern about how a subject who draws such inferences could herself understand them to be reasonable. I cannot see how Byrne's account speaks to this problem.

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

Suppose for the sake of argument that a subject concludes that she believes *p* by inferring according to BEL. Let her now ask herself what her grounds are for accepting that she believes *p*. Citing the ostensible fact that *p* sheds no light: this has no tendency to show that that she believes that *p*, as Byrne admits. What would support the subject's conclusion, of course, is the fact that she herself is ready treat *p* as a premise from which to draw inferences. But to represent *this* as her ground for accepting that she believes *p* would be, in effect, to presuppose that she already knows her own mind on the matter, and thus would undermine Byrne's account, which requires the premise of the BEL-inference to be neutral. So Byrne's approach appears to face a dilemma: either it represents the subject as drawing an inference that she should find rationally unintelligible, or else it requires her to have a kind of ground that would undermine the basic idea of the approach.[14]

Moreover – and this is my second objection – reflection on particular cases of transparency suggests that the cognitive transitions we make are not in fact transitions from *sheer* propositions about the world to propositions about the subject's own mental states. To see what I mean by saying that the relevant premises are not 'sheer' propositions about the world, consider the transition from

(1) I will ϕ

to

(2) I intend to ϕ.

I think Byrne is right that if a person judges (1), on a certain sort of basis, this also warrants her in judging (2). This would constitute a vindication for Byrne's approach, however, only if her grounds for so judging were neutral, in the sense that their availability did not presuppose an awareness of her own intentions. Now, (1) is superficially neutral: it does not refer explicitly to the subject's present mental state. But if we think carefully about the kinds of circumstances in which someone might, on the basis of thinking (1), be warranted in thinking (2), we will see that a quite special use of 'will' must be at issue here.

Let us stipulatively define a special 'intention-based' sense of 'will', 'will_{I}', whose use in joining a subject with an action-verb expresses a present intention so to act. We can distinguish 'will_{I}' from a 'will' of blank futurity ('will_{BF}'), which merely asserts that the subject will at some future time do something, leaving it open what makes this the case. In the 'will_{I}'-sense, it might be true that *I will walk to work tomorrow* (as I now intend), but false that *I will fall and break my leg tomorrow*, even if of these propositions are both true when 'will' is read in the 'will_{BF}'-sense. Now we can ask: in cases where one can move

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

transparently from (1) to (2), is the 'will' in (1) 'will$_I$' or 'will$_{BF}$'? Certainly the step is warranted if the 'will' is 'will$_I$': in this case, (2) just unpacks what the subject is already committed to in accepting that she will ϕ. But the step looks much harder to understand if her basis is simply a conviction that she will$_{BF}$ ϕ. Suppose I believe I have been subjected to hypnotic suggestion, and that as a result, when a bell rings, I will$_{BF}$ begin to cluck like a hen. It would be strange to infer from this that I now intend to cluck; and it would be no less strange, perhaps moreso, if my premonition that I will$_{BF}$ do such a thing were groundless. Yet it looks as though, in such a case, Byrne's INT-schema would encourage me to infer: I intend to cluck like a hen.

The INT-schema thus does not distinguish between convictions about my own future grounded in present intention and convictions about my future that are simply groundless. Once we recognize this, I think the INT-inference looks much less attractive. A plausible hypothesis about what grounds its initial appeal is that we are disposed to read its premise in the charitable way, as

(1a) I will$_I$ ϕ.[15]

But if my normal basis for judging that I intend to ϕ is a conviction that (1a), then although my basis is superficially neutral, it is not genuinely neutral: it presupposes an implicit awareness on my part of what I intend to do. I am, we might say, presupposing that this aspect of my future is mine to decide. Thus my basis for this self-ascription is a judgment about the world, but not a judgment *sheerly* about the world. It is, rather, a judgment about my future that presupposes something about my own cognitive relation to this future.

A plausible story about some of my beliefs about my own future is this: I hold them precisely because I *consciously intend* to make them true. These beliefs express 'practical knowledge' (or at any rate, practical conviction) about my own future: they do not involve a spectatorial or predictive attitude toward my own future, but a consciousness of this aspect of the future as mine to determine in virtue of my power to choose what I will$_I$ do. Byrne's account, however, posits the contrary order of epistemic dependence: on the basis of a sheer belief about what what I will$_{BF}$ do, I reach a conclusion about what I now intend to do. It seems to me that this would leave me with a knowledge of my own intentions that was palpably self-alienated. For such knowledge would be grounded, not in my seeing a certain act as in my power and regarding it as the thing to do, but simply in my supposing that it will in fact come to pass that I so act. Even if we grant for the sake of argument that an inference to what I intend on this sort of basis could give me knowledge that I intend to ϕ, it seems clear that it would not supply me with me an

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

agent's perspective on my future ϕ-ing: I would not be in a position to see the matter as settled *because* I so settle it.

I believe versions of these problems can be raised in every case to which Byrne applies his account. Both problems arise from the same basic feature of his approach: his resolute insistence that, when we transparently ascribe mental states to ourselves, we do not rely on any awareness of our own minds, but only on a sheer (purported) awareness of facts about the non-mental world. The result of this insistence is, first, that the subject is left without a satisfactory understanding of the rational connection between her self-ascription and its basis; and second, that her relation to the mental state she ascribes is rendered alienated and spectatorial. Even if we grant that she can know *that* she is in the certain state in this way, she does not seem to know it *from a participant's perspective*, as it were: she takes the world to be a certain way and on this basis *supposes* herself to be in a certain mental state, rather than consciously experiencing the state *in* the way she takes the world to be.

To avoid such alienation, it seems that we must reject Byrne's uncompromising insistence on a transition from sheer awareness of the non-mental world to awareness of the subject's own mental state, and allow instead that the subject's basis for her self-ascription draws on some sort of awareness of her own mind. The next view I want to consider seeks to account for transparent self-knowledge in this way.

4. Transparency as inference from judgment to belief: Peacocke

If transparent self-knowledge cannot be satisfactorily explained by an inference from a sheer proposition about the world, it seems that it must draw on some pre-existing awareness of our minds. To focus once again on the case of belief: the idea must be that, although the question on which the subject reflects is whether *p*, her transition to the self-ascription *I believe that p* depends not merely on her (seeming) awareness that *p*, but on some sort of awareness of her own consideration of the question. Christopher Peacocke (1998, 2008) has defended such a view of transparent self-ascriptions of belief. As we will see, however, the way Peacocke conceives of this approach generates a form of alienation no less problematic than the one we found in Byrne's view. Seeing this problem will sharpen our sense of what sort of self-awareness a satisfactory solution must invoke, and how this awareness must relate to the subject's 'outward-looking' awareness of the world.

According to Peacocke, a subject's ability to determine whether she believes that *p* by considering whether *p* rests on her awareness of her own act of *judgment* in response to the latter question. Judgment, Peacocke holds, is a phenomenally conscious act.[16] Hence when a subject judges that

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

p, she will be conscious, not merely of the (apparent) fact that *p*, but also of her own act of judging that *p*. This consciousness will in turn warrant her in self-ascribing a belief that *p*, since the act of judging that *p* normally expresses belief that *p* (if such a belief already exists) or produces belief that *p* (if it does not yet exist). Peacocke admits, however, that these connections do not hold universally: sometimes an act of judgment does not express or produce a standing belief. Nevertheless, he maintains, consciousness of judging that *p* is a normally reliable indicator that one believes that *p*, so when a subject has no special reason to doubt that her situation is normal, she may justifiably self-ascribe a belief that *p* on this basis. Moreover, provided her self-ascription is true, she may thereby come to know what she believes.[17]

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 sheer awareness that *p*, and holds instead that her basis is awareness of a mental event: *her judging that p*.[18] 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 non-mental world, but this 'look outward' warrants me in self-ascribing a belief only because it results in a consciousness of my own act of judgment. This indeed is how Peacocke sees the matter: he holds that his account '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, 72–3).

In spite of its naturalness, I think this account mischaracterizes something crucial about our cognitive relation to our own beliefs. We can see what is odd about it by once again considering the matter from the standpoint of a subject trying to understand her own reason for taking herself to believe that *p*. What can she say to herself?

It is essential to Peacocke's account that the step from consciously judging that *p* to knowing one's own belief that *p* be a step from one item of awareness to another, distinct item of awareness. His view, as we have seen, is that awareness of judging is a normally reliable indicator of belief that *p*, but one that can be present in the absence of such an underlying belief.[19] For my part, I am skeptical of the intelligibility of this notion of conscious judgment. After all, not just any event of consciously entertaining the content that *p* is a case of judging that *p*. If I entertain the notion of *p*'s being the case merely for the sake of argument, or in a counterfactual spirit, 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 p is true*. And it is hard to see how this can mean anything less than: it requires

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

inwardly expressing one's belief that *p*. But then it is hard to see how my consciousness of judging that *p* can 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 must include my warrant, whatever it may be, for the former

Let us suppose, however, that there is such a thing as the phenomenology of judgment: a phenomenal profile characteristic of judging that *p*, which can be present even when no corresponding belief exists. Could such phenomenology give me a reason to self-ascribe the belief that *p*? My grip on the idea is not firm enough for me to argue with confidence that it could not; but it seems to me clear that, if it did, it would at best put me in a position to have an alienated, spectatorial knowledge of my own belief. To see this, recall that 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 one consistent with its absence. On the basis of this indication, in the absence of a reason to think my situation abnormal, I am supposed to be warranted in self-ascribing the belief that *p*. So do I in fact believe *p*? 'Well,' I really ought to answer, 'it's very likely.' If I assert 'I believe that *p*' on this sort of basis, it is surely a speculation about myself, not a conscious expression of conviction that *p*. In locating my ground for self-ascribing a belief in a fallible indicator of belief, Peacocke's account drives a wedge between my consciously taking *p* to be true (which I express in judging that *p*) and my knowledge that I believe *p* (which I infer from, but am not entitled to identify with, my conscious judgment that *p*). What I can acquire by this method, if anything, is a bit of information about myself, not a consciously held stance on the question whether *p*.

If this is right, then Peacocke's account has failed to meet the constraint we drew from our discussion of Moran. For this reason, I think Peacocke's approach lost hold of the true spirit of Evans's observation, even if it fits the letter of his account. Genuinely transparent self-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-conscious stance toward the world, not an independent knowledge *of* one's holding such a stance.

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

5. The reflectivist approach

I have objected to Byrne's idea that a subject's basis for transparent self-knowledge is a *sheer* proposition about the world, and also to Peacocke's idea that her basis is a conscious event 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 knowledge of her own mind. I believe, however, that there is an insight worth preserving in each approach. In effect, 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 2011), I tried to argue that there can. The interpretation of transparent self-knowledge I proposed, which I called 'reflectivist', holds that we are warranted in self-ascribing 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 a sheer fact about the non-mental world to fact about her own psychology; nor is it, as Peacocke proposes, a transition from consciousness of one mental event to knowledge of another, distinct mental state. It is not an acquisition of new knowledge, but simply a reflective articulation of an awareness that was already involved in the subject's regarding the world in a certain way.

I continue to think that only an approach along these lines can satisfactorily explain transparent self-knowledge, but I am no longer satisfied with the account I gave of reflectivism. The crucial problem lies in making sense of the idea of tacit knowledge. For on the one hand, if a subject already *knows* (e.g.) that she believes that *p*, mustn't this involve her representing herself as believing that *p*? But then in what sense can this knowledge be 'tacit', other than the inconsequential sense that she has not yet verbalized it? And on the other hand, if the subject's knowledge that she believes *p* is truly *tacit*, in what sense can she be said already know this at all? States of knowledge are standardly thought to be individuated by their propositional objects. If a subject does not yet know 'explicitly' that she believes that *p*, what can this mean but that she does not yet possess this particular piece of knowledge? The notion of tacit knowledge thus appears to be in internal tension with itself. Yet appeal to this notion is crucial to the reflectivist position, for only in this way do reflectivists capture the distinction between the cognitive state of

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

a subject who merely has the potential to reflect on a certain belief she holds and the state of a subject who has actually done so.

The reflectivist's claim that belief involves tacit knowledge of belief is also problematic for another reason. Knowledge is commonly assumed to be subject to the following Concept Possession Requirement:

> (CPR) A subject can know that *p* only if she possesses the concepts necessary for understanding the proposition that *p*.

If the tacit knowledge invoked by reflectivism were subject to (CPR), then the reflectivist would be committed to holding that a subject can have beliefs only if she possesses the concept of belief. But this is an implausibly strong intellectual requirement on belief. The *concept* of belief seems to be a fairly sophisticated attainment, the first step toward an abstract understanding of what it is to believe, whereas it seems that quite unreflective persons, not to mention young children and nonhuman animals, can have beliefs and act intelligently on the basis of them. So it seems the reflectivist should hold that our 'tacit knowledge' of belief is not subject to (CPR). 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 does it mean to say that the subject 'tacitly knows' this?

6. Sartre and reflectivism

The reflectivist thus faces formidable challenges, but he is not alone in his predicament. Jean-Paul Sartre's distinction between 'positional' and 'non-positional' self-consciousness grew out of an attempt to respond to a structurally similar set of challenges. Some might doubt whether much help can be expected from this sort of companion, but I want to argue that Sartre's distinction is in fact of great value in responding to the problems noted above.

The positional/non-positional distinction belongs to Sartre's broader enterprise of characterizing 'consciousness', the mode of being characteristic of the entity he calls 'the for-itself'. We can introduce the distinction by explicating four fundamental propositions about consciousness asserted in the Introduction to *Being and Nothingness* (1943):

> (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) "[A]ll knowing consciousness [*conscience connaissante*] can be knowledge [*connaissance*] of nothing but its object" (BN lii/18).

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

(S3) "[A]ll positional consciousness of an object is at the same time non-positional consciousness of itself" (BN liii/19).

(S4) "[I]t 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 captures his 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 – is its intentionality, its being *of* or *about* some object distinct from the relevant state consciousness itself.[20] 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 object (i.e., that which we would specify in specifying what it is a consciousness of: an 'object' in the broadest sense). Consciousness is said to be 'positional' inasmuch as it is of or about an object.

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 he does regard knowing as a species of positional consciousness: it is a kind of relation in which consciousness can stand to a posited object. That object, and only that object, is what the relevant consciousness is knowledge *of*. This is the thought expressed in (S2).

The crucial point for our purposes is (S3): all positional consciousness of an object involves non-positional consciousness of that very state of consciousness. 'Non-positional' consciousness is supposed to be 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 a consciousness 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 is conscious *of* it?[21] Sartre argues, however, that there must be such a non-positional awareness 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 holds the that to deny the existence of non-positional consciousness of consciousness is absurd: this would be to posit 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 the notion of an 'unconscious

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

consciousness' may sound self-contradictory, but if it just consists in consciousness of an *object* without some sort of consciousness of *that very state of consciousness*, then there is really no contradiction. To insist otherwise is to beg the question. I sympathize with this response, but I think it is possible to make a more forceful case for (S3) than Sartre does here. I will turn to this task shortly. First, however, a brief remark about (S4). We have seen that Sartre holds positional consciousness of an object to depend on non-positional consciousness (of) consciousness. (S4) adds that *reflective* consciousness of our own mental states, in which we 'posit' these states as objects of knowledge in their own right, is made possible by the presence, prior to reflection, of another kind of self-awareness: a non-positional consciousness that belongs intrinsically to the relevant conscious states. Sartre holds that, when we reflect, it is this non-positional consciousness that we draw on and make explicit.

My 'reflectivist' approach resembled Sartre's view in that I also held our capacity for explicit, reflective knowledge of our own mental states to be grounded in another, more basic mode of awareness. I called this more basic awareness 'tacit knowledge', but I have come to think it better to follow Sartre in calling it 'non-positional consciousness', reserving the term 'knowledge' for an awareness that posits its object. To suggest that (e.g.) belief involves *knowledge* of belief raises all the difficulties noted earlier, and adding that the relevant knowledge is 'tacit' does not make clear how to avoid them. Characterizing the relevant awareness as 'non-positional consciousness' does not by itself provide the needed clarification, but it at least marks the spot where clarification is needed. I now want to argue that we can begin to clarify the nature of such non-positional consciousness by connecting it with the problem of transparency.

7. Non-positional consciousness and transparency

As a first illustration of how the notion of non-positional consciousness might bear on the problem of transparency, consider again transparent knowledge of one's own intentions, which we discussed earlier in connection with Byrne.

Byrne plausibly observes that a person can sometimes be warranted in answering a question about whether she intends to ϕ simply by treating it as transparent to a question about her own future: whether she will ϕ. I argued, however, that if such a transition is to give me non-alienated knowledge of my own intention, the 'will' in the question 'Will I ϕ?' must be, not '$will_{BF}$', but '$will_{I}$': a way of thinking of the future that expresses the present intention so to act. My basis for ascribing an intention to ϕ to myself is thus:

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

(1a) I will$_{I}$ ϕ.

(1a) is a specific way of thinking that I will ϕ, namely one that (as we theorists may put it) presents the relevant future action as settled by my present intention to ϕ. Nevertheless, thinking (1a) is not equivalent to thinking

(1b) I will$_{BF}$ ϕ in virtue of the fact that I now intend to ϕ.

If thinking (1a) amounted to thinking (1b), then only subjects who possessed the concept of intention and the concept of whatever relation is signified by 'in virtue of' could think (1a). But this would be an implausible intellectual requirement: surely a person may be capable of thinking about her future in a way informed by her intentions without having mastered special psychological concepts that mark her relation to the relevant future acts. Moreover – and this seems to me the deeper objection – (1b) does not capture the distinctive stance toward my future ϕ-ing that is expressed in (1a). For (1b) simply asserts a connection between two facts, one about the present (that I intend to ϕ) and another about the future (that I will$_{BF}$ ϕ). But it is clearly possible for me to think that such a connection obtains without thereby expressing the intention to *bring it about* that I ϕ expressed by (1a). (1a) does not merely posit, from a spectatorial standpoint, that a present mental state of mine will cause me to ϕ; it expresses a practical intent to ϕ. No mere claim about the relation between my present and my future, however complex, can express this sort of distinctively practical attitude unless it contains some element like 'will$_{I}$', which expresses the distinctively practical mode of the relevant thought.[22]

Thus, when I think (1a), I do not explicitly ascribe an intention to myself; rather, I think that I will ϕ in a manner that implicitly presupposes such an intention. The fact that I now intend to do this is, we might say, expressed 'non-positionally' in (1a): it is not made explicit in my thought, but it is also not something to which I am oblivious. My awareness of it will come out in the specific kinds of grounds I consider for propositions like (1a), and the specific kinds of consequences I draw from them. My grounds will speak primarily to the desirability of ϕ-ing, rather than to the evidential question whether it will be the case that I ϕ. And I will draw consequences, not about what I am *likely* to do, but about what else I must do in order to ϕ and how my ϕ-ing should affect my other plans. I will, in short, treat such propositions in ways which indicate that I understand them to express decisions rather than mere predictions. But this understanding will be expressed, not in my explicitly thinking *I intend to* ϕ, but in my distinctive way of thinking of my future ϕ-ing. I hope this example begins to clarify what, concretely, it could mean for positional consciousness of some aspect of the world (in this case,

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

an aspect of my own future) to involve non-positional consciousness of one's present state of consciousness.

Now consider a person who makes the transition from

(1a) I will$_{I}$ ϕ

to

(2) I intend to ϕ.

The reasonableness of this transition is evident. A person who soundly thinks (1a) already thinks of her future ϕ-ing in a way that implies a present intention to ϕ: her judging (2) just makes this implication explicit. What she must understand in order justifiably to make the transition from (1a) to (2) is simply that the way of thinking of her future involved in (1a) implies a present intention to ϕ. But this is to say that she does not need any further information about her present psychological state beyond what is already contained in (1a). All she needs is a grasp of this condition of application of the concept *intention* itself.

Where this is the case – where a subject's manner of thinking of the world is such that she requires only general competence with a certain psychological concept in order to know, on this basis, that she is in a certain psychological state – I will say that the subject is in a position to know her own psychological state by *reflection*. A reflective transition is not an inference from premises that are 'neutral' in Byrne's sense: accepting the relevant premises presupposes a kind of awareness of one's own psychological state, but this is a *non-positional* awareness, which does not involve application of a psychological concept. Nevertheless, such awareness can warrant a psychological self-ascription, for the application of the relevant concept just makes explicit a consciousness that was already implicit in the corresponding way of thinking of the world.

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

(3) This cat is purring.

to the reflective thought

(4) I perceive a purring cat.

How can this be a reasonable transition, given that it is one thing for a cat to be purring and another for me to perceive it? Well, consider the way the cat is represented in (3): it is presented as available for *demonstrative* reference, expressible using a 'this'. Now, an object is available to a given subject for this sort of demonstrative reference only in specific

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

kinds of circumstances. Although a 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. Philosophers commonly call such a 'this' a 'perceptual demonstrative' precisely because it expresses a mode of demonstration that is available just when the relevant object is perceived.

A subject will have the capacity to refer to objects in this way just if she is sensitive to whether she presently perceives the object in question, in a way that enables her to keep track of it and distinguish it from other objects.[23] Nevertheless, a subject who thinks (3) on the basis of perceptual consciousness does not think *that she perceives* the relevant cat: the only object she thinks about is the cat. Nor is her way of thinking of the cat is reducible to a proposition about her own mental state conjoined with a sheer proposition about her environment, as in

(3b) There is a purring cat is here and I perceive it.

(3b) does not capture the distinctively singular mode of presentation of a cat expressed by 'this cat': it represents the subject as having a nonsingular, merely existential thought about a cat that meets a certain description. By contrast, a subject who thinks (3) thinks *de re* about a certain cat, although her manner of thinking about this cat presupposes that she perceives it. We might therefore say that her perceptual relation to the cat is expressed 'non-positionally' in her thought: it is not posited, but it is a presupposition of the soundness of what is posited.

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. And if the subject goes on to think the reflective thought

(4) 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.[24] To acquire knowledge of her own perceptual state in this way, a subject need only understand the relationship between this special demonstrative mode of presentation of non-mental objects and her own perceptual state. Such a subject will thus be in a position to achieve *reflective* knowledge of her own perceptual state, provided that she grasps the general relationship between this mode of presentation and her own perceptual state. Provided she grasps this application-condition for the concept *perceives*, she will not need to draw on any further information about her present psychological state beyond what is already contained in (3). What justifies her reflective step is, however, not the sheer fact that a certain cat is purring – supposing that is so – but her *non-positional* consciousness of

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

her own manner of apprehending this fact, which is expressed in her manner of thinking of the cat.

Consider finally the case of belief. Suppose I wonder whether there will be a third world war and reach the alarming conclusion that

(5) There will be a third world war.

What I conclude here is a proposition about the non-mental world, but my manner of representing this proposition differs from the way I would represent it if I were merely supposing (5) for the sake of argument, imagining a possible world in which it held true, etc. Moreover, subjects who can deliberate competently about factual questions must have an implicit awareness of their manner of representing such propositions. For such subjects must be able to distinguish between propositions represented in the assertoric mode of belief and those represented in other modes. In particular, they must be able to distinguish between a factual question being open and its being closed: between the attitude toward *p* involved in considering *whether p* and the attitude involved in settling this question one way or another.

Now consider the kind of openness and closure that are at issue here. Suppose a person regards it as an open question whether there will be a third world war. In what sense does she regard this matter as unsettled? Not in the sense that she must regard the truth of the question as metaphysically indeterminate: she may suppose that there is a perfectly determinate fact of the matter, which she aims to discover. The sense in which she regards the question as open is rather an epistemic one: she regards it as still open *for her*, i.e., a question to which *she* possesses no determinate answer. This is not to say that she must think of herself and her own epistemic situation as such: she need only think of the proposition *p* in an interrogative mode, as it were. But her manner of thinking of this question distinguishes between a kind of openness which is in fact an openness from her epistemic standpoint and a contrasting form of closure which is in fact closure from her standpoint. When things go well, the latter mode of representation amounts to her knowing whether *p*; but whether things go well or not, representing the question as closed implies that her own *belief* on the question is settled.

The point here is not merely that the subject's answer to the question whether *p* expresses a belief she holds, but that she herself already implicitly distinguishes between this mode of representation and a contrasting non-committal mode. She might mark this distinction by using modal verbs in a way that expresses epistemic possibility, so that

(5) There *will be* a third world war

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

expresses closure of the question from her standpoint while

(6) There *might be* a third world war.

expresses openness. But however she marks it, this is a distinction that she will, as a competent deliberator, implicitly recognize. We might therefore say that in concluding that there will be a third world war, she expresses a *non-positional* consciousness of her own belief: 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 the deliberating subject then goes on to think the reflective thought

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

she will be making her attitude on the question explicit, but this awareness was already implicit in her world-directed representation of the likelihood of a third world war. To acquire reflective knowledge of her own doxastic state in this way, she need only understand the relationship between a certain mode of presentation of a worldly state of affairs and her own state of belief. And again, what justifies her reflective step will be, not the sheer thought that there will be a third world war, but her non-positional consciousness of her own stance on this question.[25]

8. Conclusion

Let me conclude by noting some advantages of this Sartrean approach to the problem of transparency.

In the first place, this approach allows us to reconcile Byrne's idea that transparent self-knowledge is grounded simply in a consideration of the world with Peacocke's thought that such knowledge must draw on some sort of awareness of our own psychological state. Sartre's idea of non-positional consciousness is the key to this reconciliation: it shows how a look outward can itself presuppose awareness of one's own psychological state without foregrounding this awareness in a way that severs the link between the subject's awareness of her own mental state and her first-order perspective on the world.

By the same token, the Sartrean approach explains why Moran is right to insist that transparent self-ascriptions of mental states express a non-spectatorial knowledge these states. For on the Sartrean view, as we have seen, transparent mental state ascriptions simply make explicit a mode of awareness that is already implicit in the corresponding outward-looking awareness of the world. Hence transparent self-knowledge cannot leave open the question whether the world is as one represents it to be; it can only be a self-conscious look outward.

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

Finally, the Sartrean approach enables us 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 Sartrean view, the consciousness 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 worldly state of affairs as obtaining. This non-positional self-awareness does not 'posit' the believer herself as a topic of knowledge or represent her condition as being of a certain kind, but provided she possesses the relevant concepts, it enables her to make a warranted self-ascriptive judgment when she reflects. 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 self-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 his distinction between non-positional self-consciousness and reflective self-knowledge is connected with another great Sartrean theme, the perpetual threat of self-alienation that characterizes our lives as conscious subjects. A major source of resistance to approaches that represent our minds as essentially self-aware, I think, is the belief that they cannot do justice to the depth and ubiquity of this threat. But precisely because it distinguishes between non-positional self-consciousness and reflective self-knowledge, Sartre's approach can readily acknowledge the many 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 in the more profound sense that it can involve a kind of 'bad faith' in which there is no stable fact to know. An adequate treatment of this theme would need to consider how our capacity for reflection is the ground both of the possibility of bad faith, and also of the conceivability of a project of authenticity. But these are topics for another essay.[26]

Notes

1. Sartre (1943), 40-41/23-24 I give page references to works by Sartre first in a standard English translation and then, after a slash mark, in a contemporary French edition (Sartre (1943), (1966). I have occasionally modified translations without comment.
2. I will assume that these ways of answering questions about our own mental states are normally ways of coming to *know* our own minds. If there are grounds for denying this, all the substantive points in this essay could be reformulated (more cumbersomely) to accommodate the point. For brevity,

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

I will sometimes just speak of 'the world' rather than 'the non-mental world'. This is simply an abbreviation: I do not deny that the world broadly conceived includes my present mental states.

3. Cf. Byrne (2011), Setiya (2012).
4. Cf. Moran (2001), Byrne (2012).
5. This question is pressed by Gallois (1996), Moran (2001), and Byrne (2018), among others.
6. The difference between knowing the content of one's attitudes and knowing that one holds the attitudes is forcefully emphasized by Dretske (2003), (2012).
7. Cf. Byrne (2005), 85, Finkelstein (2001), Postscript, and Bar-On (2004), Ch. 4.
8. Cf. Byrne (2005), 84–5 and Shah and Velleman (2005), 205–6.
9. For purposes of making my point, I will rely on the simplified understanding of the blindsight commonly discussed by philosophers. For an account of the phenomenon in its full complexity, see Weiskrantz (1986).
10. Cf. Byrne (2005), 96–8; Byrne (2011), 206–7.
11. Byrne (2012). '$[\ldots x \ldots]_V$' is supposed to be a 'v-proposition': a proposition ascribing to *x* only properties characteristically available to vision (shape, orientation, depth, color, shading, movement, etc.).
12. Byrne (2011). The INT-rule is supposed to be defeasible, and the subject must refrain from drawing it if he takes himself to believe that he will ϕ on the basis of good evidence that he will ϕ. For a related proposal, see Setiya (2012).
13. These remarks expand on points made in Boyle (2011).
14. 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 inferential disposition is reliable and safe, and then she could have a kind of second-order approval of her disposition to make the BEL inference. But 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.
15. Note that, although the English verb 'will' is ambiguous between $will_I$ and $will_{BF,}$ we do have verbal forms that strongly favor the former reading. If a person declares either 'I shall ϕ' or (more colloquially) 'I am going to ϕ', this normally expresses a present intention to ϕ. I submit that Byrne's INT-schema is initially attractive precisely because we are inclined to read the premise as tantamount to *I shall* ϕ.
16. Peacocke suggests that it is specifically an 'action awareness' (Peacocke 1998, 88, elaborated in Peacocke 2008), but this will not be crucial for the issue that concerns us.
17. Cf. Peacocke (1998), 71–3, 88–90. Peacocke goes on to acknowledge that there may be cases in which a subject self-ascribes a belief without an intervening act of conscious judgment, but he holds that, even in such cases, the subject's warrant will rest, not sheerly on her belief that *p*, but on the fact that she would have consciously judged that *p* if she had considered the question (cf. the 'requirement of first-order ratifiability' discussed at Peacocke 1998, 93–4). This complication will not matter for our purposes.

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

18. Nico Silins has defended a similar view, which he explicitly contrasts with Byrne's approach in this respect. Cf. Silins (2012), 304, fn. 12 and 306, fn. 17.
19. Cf. Silins (2012), 309, fn. 20.
20. 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. Sartre (1962), 61–8/45–51 and 109n/15n). 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.
21. To avoid outright paradox, Sartre places the 'of' in parentheses when he speaks of non-positional consciousness (of) consciousness (cf. Sartre (1956) liv/20). But this maneuver is obviously of no help without an explanation of this other mode of aboutness.
22. This is true even if we add a self-referential device to the reformulation, as in

    (1c) I will$_{BF}$ hereby ϕ because I now intend to ϕ. (cf. Setiya 2012)

    Adding 'hereby' marks the fact that my now representing this causal connection will contribute to making the relevant connection to obtain, but it is clear that even this more complex thought might express a disengaged observation about the causal relationship between various facts, rather than an intent to make things so. (On another reading, perhaps, the 'hereby' itself expresses what we have been using 'will$_I$' to express: that I resolve to make things so. But if this is what 'hereby' expresses, then it does not contribute to an account of the intention-expressing 'will$_I$' in terms of independently-intelligible materials; it is simply an alternative marker of the relevant mode of representation.)
23. Cf. Evans (1982), Ch. 6, esp. 170–176 and 192–196.
24. The presupposition of her thought may of course be false: her representation *this cat* may express a merely apparent awareness of a cat. 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 can acquire reflective knowledge of the specific sensory modality involved in a given perception, of which properties are perceived, etc. I believe all this can be done. Here I am just trying to illustrate the basic Sartrean strategy in accounting for transparent knowledge of one's own perceiving.
25. Note that, although we have developed this point with reference to an example in which a subject deliberates, the occurrence of deliberation is not essential to our account. What is crucial is that the subject's believing involves non-positional awareness of her holding a question to be closed. In a subject capable of considering propositional questions, such awareness will characterize all beliefs, even those about which the subject does not deliberate.
26. This essay was originally written for the 2013 SPAWN Conference on "Transparency of Mind" at SyracuseUniversity. I am very grateful to André Gallois for inviting me to this event and to Amy Kind for her insightful comments on the paper. The paper has remained in gestation for so long that I fear I'm not able to recall all the people who have helped me with it, but I recall particular debts to Dorit Bar-On, Alex Byrne, Matthias Haase, David Hunter, Béatrice Longuenesse, John McDowell, Dick Moran, Sarah Paul, and Kieran Setiya.

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press

Notes on contributor

Matthew Boyle is Professor of Philosophy at the University of Chicago. Before that, he was Professor of Philosophy at Harvard University. He is presently at work on a book on the self-knowledge and its importance for the philosophy of mind.

References

Bar-On, D. 2004. *Speaking My Mind*. Oxford: Oxford University Press.
Boyle, M. 2011. "Transparent Self-Knowledge." *Proceedings of the Aristotelian Society* Supplementary Volume LXXXV, 223–240. doi:10.1186/1556-276X-6-223
Byrne, A. 2005. "Introspection." *Philosophical Topics* 33 (1): 79–104. doi:10.5840/philtopics20053312.
Byrne, A. 2011. "Transparency, Belief, Intention." *Proceedings of the Aristotelian Society* Supplementary Volume LXXXV, 201–221.
Byrne, A. 2012. "Knowing What I See." In *Introspection and Consciousness*, edited by D. Smithies and D. Stoljar, 183–210. New York: Oxford University Press.
Byrne, A. 2018. *Transparency and Self-Knowledge*. Oxford: Oxford University Press.
Dretske, F. 2003. "How Do You Know You are Not a Zombie?." In *Privileged Access: Philosophical Accounts of Self-Knowledge*, edited by B. Gertler. Aldershot: Ashgate 1–13.
Dretske, F. 2012. "Awareness and Authority: Skeptical Doubts about Self-Knowledge." *In Introspection and Consciousness,* edited by D. Smithies and D. Stoljar, 49–64.New York: Oxford University Press.
Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
Finkelstein, D. 2001. *Expression and the Inner*. Cambridge, Mass: Harvard University Press.
Gallois, A. 1996. *The World Without, the Mind Within*. Cambridge: Cambridge University Press.
Moran, R. 1999. "The Authority of Self-Consciousness." *Philosophical Topics* 26 (1–2): 179–200. doi:10.5840/philtopics1999261/242.
Moran, R. 2001. *Authority and Estrangement*. Princeton: Princeton University Press.
Peacocke, C. 2008. *Truly Understood*. Oxford: Oxford University Press.
Peacocke, C. 1998. "Conscious Attitudes, Attention, and Self-Knowledge." In *Knowing Our Own Minds*, edited by C. Wright, B. Smith, and C. MacDonald. Oxford: Oxford University Press.
Sartre, J. P. 1943. *L'Être et le Néant*. Paris: Éditions Gallimard.
Sartre, J. P. 1956. *Being and Nothingness*. Trans. H. Barnes. New York: Philosophical Library.
Sartre, J. P. 1962. *The Transcendence of the Ego*. New York: Farrar: Straus and Giroux.
Sartre, J. P. 1966. *La Transcendence De l'Ego*. Paris: Librairie Philosophique J. Vrin.
Setiya, K. 2012. "Knowledge of Intention." In *Essays on Anscombe's Intention*, edited by A. Ford, J. Hornsby, and F. Stoutland, 64–97. Cambridge, Mass: Harvard University Press.
Shah, N., and J. D. Velleman. 2005. "Doxastic Deliberation." *Philosophical Review* 114 (4): 497–534. doi:10.1215/00318108-114-4-497.
Silins, N. 2012. "Judgment as a Guide to Belief." *In Smithies and Stoljar* 2012.
Weiskrantz, L. 1986. *Blindsight: A Case Study and Its Implications*. Oxford: Oxford University Press.

https://doi.org/10.1080/00455091.2019.1565621 Published online by Cambridge University Press