DOI: 10.1111/ejop.12459

BOOK REVIEW

WILEY European Journal of Philosophy

TRANSPARENCY AND SELF-KNOWLEDGE

by Alex Byrne

Oxford: Oxford University Press, 2018, xi + 227 pp. ISBN: 9780198821618. hb £30.00

The title of Alex Byrne's excellent book stakes its territory clearly: It offers an account of self-knowledge that takes the idea of transparency seriously. Helpful to the reader, but not terribly sexy. Reading it, I thought *Keep your Eyes Peeled!* might have been nice. That too captures some key themes of the book, if more obscurely. To wit: that we self-ascribe mental states with an *outward* glance rather than by introspecting conscious experience; that this is, for Byrne, the central insight behind the oft-quoted passages from Gareth Evans (1982); and, finally, that Byrne's book is something of a safari, for he is on the hunt for rules one might follow to gain knowledge of *all* of one's mental states that fit the mold, as he understands it, of belief (anyways, all the mental states that one might think we can know in a distinctively first-personal way).

One of the strengths of Byrne's book is its ambition on two fronts. First, it begins with a thorough discussion of rival theories of self-knowledge, including a detailed investigation of the various objections that have been raised against the Inner Sense account. Against what is perhaps the dominant view, Byrne thinks the account can withstand these objections, arguing that, though there are "grounds for dissatisfaction," no "knock-down refutation" has been offered (49). Second, Byrne extends his account of belief to a variety of mental states, many of which rarely if ever get discussed in the contemporary literature. In addition to belief, Byrne has fascinating things to say about knowledge, the absence of belief, perception, desire, intention, pain, emotion, memory, imagination, and thinking, much of it informed by empirical work. Byrne is interested in these cases, not just for their own sake, but because he is a *monist* about self-knowledge: he holds that there is one theory of self-knowledge that can explain our knowledge of all psychological states. It is common to reject monism, and those who embrace it tend to favor something close to the inner sense model, on which first-person access is achieved by introspecting conscious episodes. Byrne's book bucks these trends. The book is also easy and enjoyable to read, with lively prose and entertaining examples.

Because of the richness of the discussion, I will not be able to consider all of Byrne's proposals. Instead, I will focus on a few key claims, including his understanding of transparency and his extension of the model beyond belief to other intentional attitudes.

If I ask you, "do you believe the Pirates will win the pennant?", you will first answer a different question, namely, "will the Pirates win the pennant?" This is the transparency phenomenon. Many philosophers think that it captures how we actually proceed when answering questions about our beliefs and other attitudes as well. But that leaves a lot of theorizing to do, because we need an explanation of what exactly is going on. What is the transition you make here, if any? Why is the transparency procedure knowledge-conducive? Byrne offers an *inferentialist* interpretation of the procedure. On his view, one self-ascribes the belief that p by taking p as a premise in reasoning. More specifically, one achieves doxastic self-knowledge by following this epistemic rule:

BEL: If p, believe that you believe that p.

In following the rule, one infers a conclusion about one's mind from a premise about the world. Quite generally, Byrne conceives of inference in terms of conforming to epistemic rules with the form: If conditions C obtain, believe that p (e.g., "if the cows are lying down, believe that it will rain soon.").

Here are some of the advantages of Byrne's proposal. First, it has considerable intuitive plausibility, at least in that it takes the transparency phenomenon at face value. It really does seem like we form beliefs about our beliefs by first answering questions about the world. Second, it is, in his terms, "economical" (14). Our capacity for first-person knowledge is just the capacity for inference and so is already included amongt the capacities required for empirical knowledge. This contrasts with "extravagant" views, such as the inner sense approach, which hold that self-knowledge is achieved by a capacity distinct from those involved in ordinary reasoning and belief formation. Third, the view is "Detectivist" (15–16), holding that first and second-order beliefs are "distinct existences." Thus, it avoids the difficulties confronting Constitutivist approaches to self-knowledge, such as that advanced by Sydney Shoemaker (1996). (Whether this is an advantage depends on one's view of Constitutivism, of course.) Fourth, only I can gain knowledge of my beliefs by following BEL, so the proposal explains the "peculiarity" of self-knowledge—that it is achieved by a method only available to the subject. Fifth, the proposal explains epistemic privilege. Suppose that p is false. Still, if I follow BEL, then I will be guaranteed to form a true second-order belief. Thus, following BEL is a "self-verifying" procedure (104–105). This is because recognizing that p is the manifestation or onset of belief. Indeed, BEL is "strongly self-verifying," because one will arrive at a true second-order belief if one so much as *tries* to follow it (107). (On Byrne's reckoning, one tries to follow a rule if one believes that the conditions mentioned in the antecedent obtain.) Thus, on Byrne's view of belief, the state known is itself involved in the process of coming to know it. You know what you believe thanks to manifesting or forming that belief. This is a compelling idea, especially for those impressed by Evans' thought that we form self-ascriptions by putting into operation the same capacities involved in forming the first-order states themselves (1982, 227).

The most obvious difficulty with this proposal, which Byrne calls "the puzzle of transparency," is simple: p is poor evidence that I believe that p. His response is to reject the demand that beliefs be based on good evidence. Instead, BEL is knowledge-conducive because beliefs formed by following it are safe. What the account shows is that there are cases in which "knowledge can be obtained by reasoning from inadequate evidence, or from no evidence at all" (208). There are reasons to be unhappy with this solution to the puzzle, of course. Many have been raised in the literature already, and Byrne includes helpful discussion in the book (121–127). Instead of festering on the case of belief, I'll shift to Byrne's project of extension. Byrne's account of how we know our beliefs is appealing. Are the extensions?

There is room for doubt on this front because some of the appealing features of the account of belief do not apply in the other cases. According to Byrne, here are the rules we follow in order to know our desires and feelings of disgust:

DES: If Φ-ing is a desirable option, believe that you want to Φ. (161)

DIS: If x is disgusting, and produces disgust reactions in you, believe you feel disgust at x. (178)

As should be clear, neither rule is self-verifying. Instead, Byrne notes, the rules are defeasible (161, 164–166). Still, they are "practically self-verifying" (162). This is because, in "all ordinary situations" (140), one cannot know the premise unless one is in the mental state one goes on to ascribe. And furthermore, these are good rules to follow even if the premise is false. Even if drinking the glass on the table is not desirable, I can truly self-ascribe desire from the premise that it is. This depends on the plausible thought that there is a reliable connection between what we judge to be desirable, or disgusting, and what we actually desire or feel disgust about. (As Byrne puts it: "known desirable options tend to be desired" (161)). A second difference from BEL is that the mental states of desire and disgust are not themselves manifest in the psychological processes of following the rules for self-ascribing them. One forms a self-ascription on the basis of a consideration that is reliably connected to these attitudes, as opposed to their actual manifestation. More contentiously, it is far from clear that these models have the intuitive plausibility that his approach to belief does. Evans' original observation accurately captures what we actually do when we form beliefs about our beliefs. DIS, at least to me, does not. Attention to the object of my emotion seems to play a role in knowing

how I feel, but first, it does not seem like attention to my own behavior is required at all, nor does attention to the object seem to suffice. The most natural view, one would think, is that I know how I feel on the basis of the character of my experience.

Setting the phenomenology aside, one might ask whether the other differences are significant. I think they are. It is not clear that a rule that is merely practically self-verifying is a genuinely peculiar method for forming beliefs about my mind. As mentioned, DES is a good rule to follow on the assumption that there is a reliable enough connection between my judgments about what is desirable and my desires. But why can't there be a reliable connection between another person's judgments about what is desirable and my desires? Consider Fred and Barney, two inseparable friends who share all of the same hobbies, desires, and tastes. Barney can form beliefs about Fred's desires by following this rule:

FDES: If Φ-ing is a desirable option, believe that Fred wants to Φ.

FDES is not self-verifying; it is defeasible. But so is DES. Suppose that Fred follows DES. Fred's method is basically the same as Barney's, so what is first-personal about Fred's? Notice that the fact that Fred's method works even if the premise is false does not help: That's true of FDES, too. To be clear, the worry is not that DES is a bad rule to follow or that it is not knowledge-conducive. The issue is whether following DES yields genuinely *first-personal* knowledge.

Byrne considers an analogue of BDES for belief (109).

BEL-3: If p, believe that Fred believes that p.

According to Byrne, BEL-3 is a bad rule to follow because: "Science fiction aside, your belief about Fred is responsive to what *you* believe, not what *Fred* believes" (ibid.). It's not clear why we can set science fiction aside, though. If peculiarity means that, in principle, introspection is a method only available to the subject, then it should not be conceivable that another uses it. Regardless, it is plausible that BEL captures the peculiarity of self-knowledge *because* it is strongly self-verifying. And it is strongly self-verifying because the mental state ascribed is manifest in the process of ascribing it. Only the subject herself can form a self-ascription of belief *that* way. But DES and DIS are different. When you follow DES, your beliefs about your desires are *not* based on your desires. They are based on your judgments about what is desirable. Those judgments, in turn, are responsive to what you desire. But Barney's judgments are responsive to what Fred desires, as well. So there does not seem to be a significant difference here.

Byrne might respond by holding that there is a constitutive connection between one's desires and one's judgments of desirability. (Officially, Byrne is neutral about whether the connection is contingent or constitutive (161)). That would distinguish DES and FDES. But it's not clear that Byrne can appeal to this, since, for him, the rules in question are licensed only by the safety of the beliefs arrived at by following them. The worry is that once we shift to rules that are merely practically self-verifying, it is too easy to come up with "transparent" rules that would yield safe beliefs about another's attitudes. We only need people who are similar in what they are into. If this is science fiction, it is not terribly extravagant.

Another worry I had was whether DES and DIS are genuinely *transparent*, even setting aside the premise about one's own reaction in DIS. Byrne's talk of self-verification misses out on something one might think is important about the case of belief: when I answer the question about the world that simply *settles* the question about my mind. Once I have concluded that p there is nothing for me to think other than that I believe that p. This is not the same thing as the claim that following a rule is guaranteed to arrive at true beliefs. It is a claim about how things seem from my point of view once I have looked outward, even before I make the transition to a self-ascription. But this feature is not found in the other cases. Judging it desirable to φ does not settle the question whether I desire it. Nor does judging x disgusting settle the question whether I am disgusted by it, even if I add to it a judgment about my own behavior. It is intelligible to think: "It is desirable to φ, but do I want to?" but not: "p, but do I believe that p?" Again, it is plausible to think that the difference is explained, at least in part, by the fact that judging p is a manifestation of the belief that p whereas judging something desirable is not a manifestation of desire. If you think that having a question about one's mind *settled* by a question about the world is a mark of transparency, then you should worry about extending the model as Byrne does.

Byrne's book helpfully brings to the surface an important question, namely, what makes a method for forming self-ascriptions transparent? As I understand it, Byrne's answer is *negative*: A transparent procedure is one that does not avail itself of the deliverances of an "inner glance" at the phenomenal character of one's experience. So it is no great problem for him that some features found in the case of belief are absent in others. But others will not be persuaded, because they favor a *positive* conception of transparency. For example, they might require that the first-order mental state itself plays a role in the procedure for self-ascribing it. Or they might insist that if answering a question about the world does not settle a question about a mental state, then that state is not transparent, even if an "outward glance" plays some part in the epistemology.

Many philosophers think that the transparency phenomenon should be at the center of our attempts to understand self-knowledge of attitudes like belief. And many philosophers, regardless of their stance on transparency, share Byrne's monist ambitions. Byrne's book is the most thoroughgoing attempt to offer a monist transparency theory. I have raised some worries about what gets lost when the model is extended to other attitudes. Perhaps there is reason to favor a positive over a negative account of transparency. However, the more demanding the notion of transparency, the less likely it is to generalize, and so the dimmer the prospects for monism. On the other hand, if monism is false, does that mean that talk of "the first-person perspective" is ambiguous? These issues are at the heart of contemporary debates about self-knowledge. Anyone with an interest in these debates should read Byrne's book.

Casey Doyle
St Hilda's College, Oxford, UK

WORKS CITED

Evans, G. (1982). In J. McDowell (Ed.), *The varieties of reference*. Oxford: Clarendon Press.

Shoemaker, S. (1996). *The first-person perspective and other essays*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511624674