Review Essay

EYES DIRECTED OUTWARDS

TRANSPARENCY AND SELF-KNOWLEDGE, Alex *Byrne* Oxford: Oxford University Press, 2018, 227pp., $40.95 (hbk), ISBN 9780198821618.

Paul Conlan, Giovanni Merlo, and Crispin Wright

I

In an oft-cited passage in *The Varieties of Reference*,[1] Gareth Evans wrote that

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

Evans' suggestion is that, at least in a significant range of cases, a question about one's doxastic state is, in a now widespread terminology, *transparent to* a question about the non-mental world: answer the latter to your satisfaction and you thereby answer the former. Indeed, not only *can* you answer the psychological question in this 'outward-looking' kind of way, but moreover, Evans seems to suggest, you *must* do so.

That last claim seems clearly wrong. If you are asked, 'Do you think there is going to be a third world war?', you might perfectly (self-) knowledgeably answer, without "directing your eyes outward", that you have never thought about the matter and have no view. If, it may be said, Evans is right to suggest that, in a normal context, you would have to answer the 'Do you think. . . ?' question in the same way as would be appropriate for the 'Will there be . . ?' question, this is because in a normal context, there would be no interest in the distinction between the questions — the 'Do you think . . .?' question would not be heard as a psychological question at all.

It may be rejoined that Evans mis-stated a good point: that his crucial insight was not that one *must* answer the 'Do you think . . .?' question by addressing the worldly one, but that one *can* so answer the former, even when it is taken strictly psychologically. You can settle a question about your mind by settling a question about non-mental reality. This 'transparency' thesis has been widely received as of fundamental importance for a satisfactory account of the

[1] Evans (1982). Henceforward *Varieties*.

apparent immediacy of much ordinary self-knowledge and of the special authority of first-person reports of attitudes and other states of mind.[2]

Now it's notable that there is no suggestion in Evans himself that this 'transparency thought' even provides the key to a workable overall account of self-knowledge of belief in particular, still less that it provides a pointer towards a satisfactory general account of psychological self-knowledge of other attitudes, or of the wider field of psychological states *tout court*.[3] In this trenchantly and engagingly written book (hereafter: TSK), however, Alex Byrne goes the whole hog, elaborating and defending a form of transparency theory of psychological self-knowledge for each of belief, perception and sensation, desire, intention, emotion, memory, imagination and thought. Specifically, he argues that for mental states and events of all these kinds, one comes to know that one is a particular such state by *inference* — as we shall say, a *Byrnean* inference —from a worldly or environmental premise to the conclusion that one is in the state in question. Note that the 'transparency' entrained by such a view is really that of the *non*-mental environment: it is this that one supposedly 'sees through'. Self-knowledge is canonically achieved by inference from materials delivered by attention to a suitably corresponding tract of the non-mental world, not by any kind of direct scrutiny of what lies within the mind.

There are two immediate misgivings about this very striking project. First, any such account for a particular mental characteristic, M, must of course first propose some suitable tract of non-mental reality attention to which is to provide the basis for the relevant Byrnean inference to the obtaining of an instance of M. That one might accomplish this for the general run of mental states other than belief impresses as far-fetched. What, for example, is to be the worldly premise so corresponding to 'I intend to do this', or so corresponding to 'I am suffering a visual aurora'? Second, while belief is doubtless the most opportune case, where the relevant form of Byrnean inference can, as prefigured in Evans, simply be that from 'P' to 'I believe that P', and where someone who recognizes the premise as true and moves to the conclusion will, plausibly[4] conclude something true, the transition, as Byrne is only too well aware, *qua inference*, seems epistemically terrible:[5] even the most unsophisticated self-

---

[2] It is, for example, central in Richard Moran's treatment of self-knowledge in his (2001), where the central contention is that questions about one's attitudes are characteristically resolved by attending to what they *ought* to be, and in Jordi Fernandez' (2013) which undertakes a systematic defense of the view that we attribute beliefs and desires to ourselves based on our *grounds* for those beliefs and desires.

[3] In *Varieties*, Evans does advocate an 'outward-looking' account of self-knowledge of one's occurrent perceivings (see pp. 227-228), but, in introducing this account, he is careful to describe it as different from the one he proposed for self-knowledge of one's beliefs (p. 226). The basic idea is that, in determining what he is perceiving, a subject will "[go] through exactly the same procedure as he would go through if he were trying to make a judgement about how it is at this place now, but excluding any knowledge he has *of an extraneous kind*" (p. 227).

[4] We say "plausibly" rather than "necessarily" because of the arguable additional dispositional strength of 'X believes that P' over 'X judges that P'. Byrne, on the other hand, regards the inference as self-verifying and the truth of its conclusion as guaranteed. We'll come back to this below.

[5] "Mad" according to Matthew Boyle (Boyle, 2011, pp. 7-8)

knower will surely be aware that the obtaining of the circumstance depicted by 'P' is in general at best exiguous evidence that one believes that P.

In the later chapters of his book, Byrne cheerfully takes these challenges on. But first he offers the reader a systematic and clear overview of the state of the contemporary debate, as he sees it. In chapter 1, he presents the problem of self-knowledge as that of explaining the 'privileged' and 'peculiar' access we enjoy to our own mental states.[6] Chapter 2 discusses *inner sense* accounts of this problem, corralling a series of objections to such accounts, responses to those objections, and residual difficulties that remain after these objections have been answered. Chapter 3 surveys three recent much-discussed alternative approaches due to Davidson (1984), Moran (2001) and Bar-On (2004), respectively, raising objections to each of them. Chapter 4 examines a problem for the transparency thought itself — what Byrne calls the 'puzzle of transparency'. Chapter 5 offers a solution to the puzzle of transparency for belief. The remainder of the book proceeds to attempt to develop accounts of the general Byrnean-inferential form covering each of perception and sensation (chapter 6), desire, intention, and emotion (chapter 7), and memory, imagination, and thought (chapter 8).

Byrne explains that for each one of these cases he seeks an account that meets the following four constraints.

First, in contrast to the so-called Constitutivist proposals of Wright (1987), Bilgrami (2012), Coliva (2016; 2017) and others, a satisfactory account must be uniformly *detectivist*. Byrne writes

> Detectivist accounts liken self-knowledge to ordinary empirical knowledge in the following two abstract respects. First, causal mechanisms play an essential role in the acquisition of such knowledge, linking one's knowledge with its subject matter. Second, the known facts are not dependent in any exciting sense on the availability of methods for detecting them, or on the knowledge of them—in particular, they could have obtained forever unknown. (TSK, p. 15)

The states of affairs which psychological self-knowledge is characteristically about are thus to be conceived as self-standing, and in basic cases, independent of their recognition by the subject.

Second, as indicated, the accounts of the various kinds of self-knowledge addressed are to be uniformly *inferentialist*, with the direction of inference always from non-mental world to mind.

Third, a satisfactory account must be *economical:* the cognitive resources which it represents us as deploying in achieving routine self-knowledge of any of the relevant kinds should be restricted to capacities which are involved in the achievement of knowledge of other non-psychological kinds. So, in particular, no special mechanisms whose distinguishing task it is to track the mental may be admissibly invoked. This repudiation differentiates Byrne's account from the *Monitoring Mechanism* account of Nichols and Stich (2003), and Armstrong's *Self-Scanning* account (Armstrong 1968), both of which posit a special detective capacity supposedly exclusively deployed in the achievement of self-knowledge .

---

[6] We will explain Byrne's understanding of these terms shortly.

Fourth, the account is to explain both the relative epistemic security of beliefs about one's own mental states compared to that of one's beliefs about others' mental states —"*privileged access*" in Byrne's interpretation of that phrase— as well as the fact —"*peculiar access*" is Byrne's term — that one seemingly has a special first-personal way of knowing about one's own mental states which contrasts with the ways one can know about the mental states of others.

Each of these constraints may be challenged. But in our view much previous work on self-knowledge has suffered from want of the kind of discipline they collectively exert. Should it prove impossible to satisfy them, the reasons why may be expected further to illuminate the philosophical subject matter. In the remainder of this review essay, we shall concentrate—partly for reasons of space, but also because, Byrne's optimism and energetic efforts notwithstanding, there is no evident reason why a transparency account should be expected to succeed across the board,— on whether he manages to fashion a satisfactory account complying with all four of his constraints at least for the marquee, and perhaps most promising, case: that of belief.[7]

II

Here, exactly as the reader will expect, Byrne's core contention is that our self-knowledge — or at least, that part of it that impresses as manifesting privileged and peculiar access —is achieved by inference in accordance with the rule:

BEL If P, believe that you believe that P

It is important to stress that Byrne's claim is that we *actually do follow* this rule—an account is aimed at of how we do actually routinely know of our beliefs, (not a mere model of how we might.)

There are three separable salient kinds of misgiving concerning this proposal.

The first concerns the *nature of the transitions* across BEL that, in Byrne's view, we habitually make and the *epistemic standing* of the results. If a thinker follows BEL as a rule of inference and forms thereby a true belief that she believes that P, what makes that belief *knowledgeable*? Byrne makes a case that the procedure is safe because *self-verifying* — indeed guaranteed of safety even if one botches the investigation and hasn't really verified P at all.[8] However, safety

[7] The Appendix to this essay offers some reflections on some of the details of his more general project, case by case.

[8] See TSK, pp. 109-112. In Byrne's terminology, BEL is *self-verifying* because "if it is followed, then the resulting second-order belief is true" (TSK, p. 104) and *strongly self-verifying* because "if one *tries* to follow it, one's second-order belief is true." (TSK, p. 107). The difference between following and trying to follow BEL lies in whether one's 'outward gaze' delivers knowledge of the tract of non-mental reality on which it is directed: when one follows BEL, one forms the belief that one believes that P because one *recognizes* (hence, knows) that P (TSK, p. 101-102), whereas when one tries to follow BEL, one forms the belief that one believes that P because one *believes* that P (TSK, p. 107).

is not, plausibly, sufficient for knowledge, at least when the relevant belief is acquired by inference. A subject who arrives at a true belief by inference had better, if that belief is to count as knowledgeable, also be *rational* both in making and in accepting the conclusion of the inference — or at least not irrational. But must not the agent of a Byrnean inference across BEL be regarded as egregiously irrational if they are thereby led to *base* their acceptance of its conclusion on its premise? And while the nature of inference, *qua* inference, and of the basing relation are, to be sure, stubbornly unsettled topics in contemporary epistemology, it is very plausible that they belong together at least to the extent that inference, whatever else it might be, is essentially a movement of thought which, in cases when it transitions from a belief to a new belief, formed as a result of the inference itself, essentially initiates a basing relation between the former and the latter.

If that is right and if knowledgeable belief, even if safely formed, cannot, *qua* knowledgeable, be belief held on the basis of irrelevant— 'terrible' or 'mad'— reasons, then there is a case for saying that beliefs arrived at by inference across BEL are not, as such, (self-) *knowledge*. That verdict, would, of course, completely deflate Byrne's project.

At this point, it may be wondered whether Byrne is not perhaps misplaying the hand that Evans has dealt him. Perhaps he undersells the real credentials of BEL by invoking the idea of inference at all. Might he instead have proposed, to better effect, that BEL and its kind encode not rules of inference but merely *patterns of transition* which we (unreflectively) conform to in forming beliefs about our own mental states? Is there any advantage of the Byrne-inferential account that would be jeopardized by such a shift? As the reader will speedily appreciate, Byrne's explanations neither of privileged access (TSK pp. 109-12) nor of peculiar access (TSK pp. 108-9) would need to proceed differently on such a proposal.[9] Admittedly, Byrne seems heavily invested in a view of Byrnean transition that is flatly at odds with this.[10] But does the best implementation of his basic Evansian idea require him to be so?

---

[9] Simply: that transitions in accordance with BEL are, if Byrne is right, strongly self-verifying suffices to ensure privileged access as he understands it, —that selves' beliefs about their own beliefs are relatively epistemically secure compared to their beliefs about the beliefs of others,— while peculiar access is ensured by the point that transitions across BEL only work in general to secure true second-order beliefs in cases in which one self-applies the method. Nothing would be lost then, at least in those respects, if Byrne were to drop the notion of inference. The reader may further reflect that a denial that transitions in accordance with BEL should properly be accounted as inferences would surely be enforced if, with Paul Boghossian, it is accepted as constitutive of inference in cases where it is a movement from beliefs to new beliefs that the subject "take" the former to support the latter (Boghossian 2014), since no-one is going clear-headedly to take the premise for a Byrnean 'inference' across BEL to provide such support for the 'conclusion'.

[10] He writes, for example, (at p. 101):

> So, what does it mean to say that Mrs. Hudson *follows* this rule on a particular occasion? Let us stipulate, not unnaturally, that she follows the rule just in case she believes that there is someone at the door *because* she recognizes that the doorbell is ringing. The 'because' is intended to mark the kind of reason-giving causal connection that is often discussed under the rubric of 'the basing relation.' (TSK, p. 101)

The trouble with this rescue is that it would arguably forfeit the constraint of Economy. For what could explain our conformity to such patterns of belief-formation but the operations of a special detective mechanism—something which, given e.g. the input of a cognitive state encompassing a subject's recognition that P is true, would simply cause them to form the corresponding second-order belief? This need not be, to be sure, exactly the usual 'special mechanism' type of account, whereby it is the first-order state that is alleged to include causation of the corresponding second-order state in its essential functional role. But — whether or not one buys Byrne's view of Economy as a general theoretical desideratum — this variant appeal to a special mechanism would certainly inherit the characteristic drawbacks of its more orthodox relative.[11]

However, it may occur to a sympathetic reader that there is, perhaps no need for Byrne to run this risk. He has an intermediate option, so to say: that of representing BEL as more than a ~~(sub-personal)~~ pattern of brute transition but as less than a rule of inference. He can propose that its status is that of a rule (of belief-formation) which we *follow*, but not as a rule of *inference*. This strategy could still be economical: the epistemic capacities in which self-knowledge of belief would be grounded would be whatever would be involved in recognizing the truth of a (suitably non-mental) antecedent of BEL plus whatever is involved in following a general conditional rule. Provided that following BEL can be distanced from the idea of following a rule of inference, but is to be viewed, rather, as the following of another kind of belief-forming rule not beholden to the same epistemic standards as inference, Byrne could thereby finesse the concerns

---

These remarks occur in the context of a specific toy epistemic rule DOORBELL ('If the doorbell rings, believe that there is someone at the door') but it is clear from his proposal that a connection with basing is to be a feature of Byrne's conception of his epistemic rules in general.

[11] Byrne offers relatively little by way of explicit argumentation for the constraint of Economy, merely expressing a moderate skepticism about the extent to which philosophers are entitled to suggest un-economical accounts of privileged access. He writes that "there would appear to be little that the philosophical proponents of inner sense can contribute to our understanding of self-knowledge beyond a few pages motivating their theory, and some general discussion of the epistemology of perception" (TSK, p. 115, fn. 17). It seems to us that he undersells his case. At any rate, there are powerful reasons to abjure philosophical approaches to self-knowledge that postulate special detective mechanisms. What, after all, can a mere philosopher hope to contribute to the explanation of privileged and peculiar access? It cannot be determined from the armchair how privileged and peculiar self-knowledge is actually, as a matter of contingent empirical fact, so pervasively and apparently effortlessly achieved. What the philosopher can reasonably hope to do is to articulate one or more plausible models of how, as a hypothesis, it is achieved. But such a model, if it is not to seem entirely *ad hoc*, must necessarily proceed in terms of cognitive capacities which we already know we have and which impress as relatively well-understood. They will therefore perforce be capacities which we take to be exercised in other domains. Moreover, insofar as the kind of special mechanism in question is conceived of as involving a (sub-)personal-level 'scanning' process whereby the subject detects his or her own mental states, the resulting kind of account must, it would seem, leave room for the possibility of "brute" errors; that is, in Burge's terminology, a kind of "error that indicates no rational failure and no malfunction in the mistaken individual" (Burge, 1996, 101), —and so must confront strong arguments (Burge, op. cit.) that such errors seem not in general to be possible where self-knowledge is concerned.

we registered about basing and irrational inference, while preserving Economy. He might add that the reliability of the transitions sanctioned by BEL is enough to guarantee the epistemic credentials of the products of BEL without it being considered as a rule of inference.

To stress, this is not Byrne's view. But it too would have discomforts. One concern is that it is not clear what it would take for us to 'follow' a rule when the rule in question conditionally prescribes action which is beyond our voluntary control – as, arguably, in the case of forming a belief. It may be countered that this is a worry for the notion of epistemic rules *per se*, insofar as they purport to control belief-formation and revision, and hence is nothing new: that it was already a shadow over the Byrnean rules when conceived as originally by Byrne, as involving inference and basing, for then, too, the output is beyond voluntary control. But there is a strong reason why the concern may seem more pressing when, as on the 'intermediate' proposal, inference and basing are out of the picture. To see the point, consider the directive to a jury, 'If on balance the evidence seems to you to be such as to clearly incriminate the defendant, then come to the opinion that they are guilty.' It may be suggested that a rational agent intending to follow this rule, will encounter no analogue of the kind of potential lacuna which, by contrast, would enter into following of the rule, 'If on balance the evidence seems to you to be such as to clearly incriminate the defendant, then return the verdict that they are guilty.' Free agency is implicated in the following of the second. Despite the weight of the evidence, you might choose to conspire with the other jurors to acquit. But arguably free (intellectual) agency is not involved, for a rational subject, in following the first rule: in responding to an appreciation of the satisfaction of its antecedent in the way required by the first rule, one's doxastic response to what one regards as the balance of evidence will, unless akratic, be delivered automatically.

In short: the whole idea that we do indeed routinely follow epistemic rules of belief-formation passes muster, unless one subscribes to a full-blown voluntarism about belief, only if an appreciation that their antecedent conditions obtain is such as normally to generate the type of belief in question in a rational subject without the need for any kind of supplementary volition. There is accordingly a doubt whether, if Byrnean rules were to be proposed as having the status postulated by the 'intermediate' suggestion—as non-inferential rules of belief-formation, demanding merely that when certain conditions obtain, one is to form a certain belief, albeit one on which those conditions need have no evidential bearing, —it would make good philosophical-psychological sense, to suppose one might be capable of following them at all.

And so each of three possible construals of Byrnean rules—as rules of inference, as patterns of brute transition, and as non-inferential rules of belief-formation, —seems beset with objection and difficulty.

III

We announced three groups of misgivings concerning the prospects for a satisfactory epistemology of self-knowledge of belief based on BEL. The second is a cluster of worries about the apparent limitations of the scope of BEL: about whether there is any clear prospect of accounting for the totality thereby of our seemingly privileged and peculiar self-knowledge of belief and cognate doxastic

attitudes. Even if it is granted that BEL is a second-order belief-determining rule that we often deploy, rationally arriving at knowledgeable beliefs thereby, it's prima facie subject to at least three kinds of limitation which suggest it would be overly sanguine to expect that a comprehensive account of self-knowledge of doxastic states might be based upon it.

First, there is arguably a difference between belief proper — a disposition-like state — and the episodic state of *judging*: that is, of coming to a view as a specific, datable event. The reliability of BEL for which Byrne argues properly pertains to the latter. But to judge that P in the latter sense is no guarantee that one acquires the relevant disposition. The judgement may not 'stick'. This point opens a lacuna between an episode of following BEL and the formation of an appropriate belief, dispositionally understood.

Second, a common objection to transparency accounts of belief is that in a wide class of cases, if the question whether one believes that P is put, one finds one's mind *already* made up, without necessarily any recollection of how came to be so. There is no need in such a case to 'direct one's eyes outwards' in order to answer the question. One is nevertheless able accurately and authoritatively to report one's belief, without any apparent reliance on reasons or interpretation.

Byrne has responded to this concern in conversation[12] by suggesting that, when one's mind is already made up, one simply remembers that P —in the so-termed *semantic*, rather than *episodic* sense of 'remembers', as in 'X remembers the seven-times multiplication table' or 'X remembers that the French Revolution started in 1789'.[13] So, such cases are still applications of BEL: it is just that the method of investigation called on for the relevant antecedent of the rule is that of an exercise of semantic memory. The transition is still from a (semantically remembered) worldly fact to a second-order belief.

This response, it may be countered, will not cover all cases. Consider a subject's neurotic, obsessional belief that there will indeed be a Third World War, or a mother's conviction of the innocence of her son, caught red-handed stealing. In such cases there need have been no forgotten investigative process whose result is then semantically remembered. Indeed, there need have been no investigative process at all. Yet the groundless beliefs involved may still be reportable with the usual characteristic authority and immediacy.

It may be replied that such cases are perfectly consistent with second-order belief-formation via BEL. We ask the neurotic, 'Do you believe that there will be a Third World War?', he asks himself, 'Will there be a Third World War?"', answers in the affirmative, and, transitioning via BEL, reports that he does indeed have that belief. But this way with the objection is uneasy, since it plays down what was supposed to be the principal attraction of the whole transparency direction, namely that one settles a psychological question by attending to a non-psychological tract of reality. For what such tract exactly is the neurotic attending to? Sure, the bluff answer is available: 'The matter of whether there will be a third world war'. But the truth is that what he is really attending to is his preformed opinion on the matter, and not on worldly matters

[12] In the book Byrne attributes this response to Moran (TSK p. 61), but the context seems to confirm that he means to endorse it.

[13] For the distinction between semantic and episodic memory, see Squire (1992, 232-3).

that bear on the truth of that opinion. So the suggestion that his original answer is delivered via an application of BEL is, in the intended spirit of the transparency thought, arguably a sham.

Third, the upshot of an investigation whether P can be any of a whole plethora of doxastic attitudes besides formation of a belief that P (or for that matter a belief that not-P.) One may fail to come to a view, or indeed decide that there is no justifiably coming to any view, or arrive at the belief that P is on balance more probable than not but nevertheless withhold commitment to P,… or any of a number of other possible doxastic upshots. In all these cases one would be expected to be able to pronounce authoritatively about one's resultant attitude. Recognising that BEL gives us no grip on this capacity of nuanced second-order doxastic discrimination, Byrne moves (TSK pp. 118-121) to supplement his account with two additional epistemic rules:

> NOVIEW If you are in a poor epistemic position as to whether P, believe that you do not believe that P.
>
> CONFIDENCE If you believe that P, and your belief has high (low) epistemic credentials, believe that you believe that P with high (low) confidence.

Clearly, ordinary subjects do not think of themselves *as* being in a "poor epistemic situation" or of their beliefs *as* having "high (low) epistemic credentials", —these concepts are the stock-in-trade of epistemologists —but Byrne suggests that "these are just schematic terms, to be filled in with what may well be a grab-bag of cues and heuristics, varying from occasion to occasion"(TSK, p. 120). For example, my knowledge that I am in a poor epistemic situation as to whether it's raining in New Delhi may be cashed out as knowledge that "I am currently nowhere near New Delhi, I have not read *The Times of India* today, or spoken by telephone with anyone living in New Delhi" (TSK, p. 118). The problem, though, is that, 'cues' of this sort provide, at best, defeasible evidence for the second-order beliefs they are meant, via inference across NOVIEW and CONFIDENCE, to support – so, unlike BEL, neither of those rules is guaranteed to generate true second-order beliefs.[14] This means that, if Byrne's proposal is correct, a belief that one believes that P and a belief that one *doesn't* believe that P (or that one believes that P with high (low) confidence) will *not*, when formed in the normal way, enjoy the same degree of epistemic security – the latter will be radically less secure than the former. And that goes against what seems, at least initially, plausible: that self-knowledge of unbelief and of

[14] As Byrne himself acknowledges (TSK, p. 118), there is also the problem that one's access to the relevant 'cues' is not particularly privileged: I may *falsely* believe that I am currently nowhere near New Delhi and that I have not read *The Times of India* today, or spoken by telephone with anyone living in New Delhi, in which case, following NOVIEW, I will form a *false* belief that I don't believe that it's raining in New Dehli.

how confident one is in a given proposition, is epistemically on all fours with self-knowledge of belief.[15]

IV

There is finally, if correct, a lethal objection to the entire approach. *We do not actually follow BEL.* Recall, as was emphasized earlier, that Byrne is aiming to give an account of the *actual* provenance of our self-knowledge. Now, we can be said to have a practice of following a conditional rule just to the extent that, provided the antecedent obtains, we make some significant effort to comply with the consequent, or at least regard ourselves as in some kind of default if we do not. But we do nothing of the sort with BEL: each of us is content to put up with (the realization that there are) no end of cases where a proposition P is true, and yet we form no second-order belief. There are no end of true propositions about which we have no second-order beliefs, and have no concern whatever to arrive at any.

To be clear: this is not a matter of 'clutter-avoidance', in Harman's (1986) sense of the term: the point is not that it would merely be impractical to 'store' all the beliefs that BEL would have us form, but that we are perfectly content to form infinitely fewer beliefs than – if we really had a practice of following BEL – we would be required to form. For example, Proxima Centauri either has more than two planets or it does not. If it does then, by BEL, we should form a belief that we believe that it does. Likewise if it does not, we should believe that we believe that. One way or the other then, if we are BEL-followers, there is a second-order belief we should have about the satellites of Proxima Centauri. But not only do we not have any such second-order belief: we think that, in our present evidential situation, we ought to believe neither hypothesis about Proxima Centauri, and hence, if we have managed our beliefs as we ought, *are right* to lack each of the second-order beliefs which compliance with BEL would, one way or the other, require.

Now, this is liable to seem a daft objection. Some readers may want to respond to it by suggesting that it simply perversely misconstrues the gist of BEL; specifically, that the rule is, in intended effect, that when enquiry leads you to the conclusion that P, believe that you believe that P.[16] But any such refinement of the intent of BEL would obviously be completely antithetical to Byrne's purpose, since recognition of when one's enquiry has led one to the conclusion that P will invariably incorporate in effect a substantial package of attitudinal self-knowledge, encompassing awareness of a range of collateral beliefs and of physical and/or intellectual actions one has performed. To arrive at a knowledgeable second-order belief in *that* way is to transition not from without but from *within* the domain of self-knowledge. So the suggestion compromises the basic pre-requisite for a Byrnean transition, viz. a premise or trigger proposition concerning non-mental reality.

---

[15] As we explain in the Appendix, Byrne's account of self-knowledge of mental states other than belief predicts the same prima facie implausible disparity. The reader should be aware, though, that, at least in the case of NOVIEW, Byrne regards this as a welcome result (TSK, p. 118).

[16] Of course, to gloss the rule in this way is immediately in tension with the response, canvassed above, to the examples of the besotted mother ad the neurotic.

Actually, we need a distinction here. It is opportune at this point to refine our understanding somewhat of the rules of engagement under which Byrne is or ought to be operating. It won't suffice for his purpose if the antecedent of a conditional Byrnean rule, proposed to be operative in our self-knowledge of a certain kind of mental state, merely *overtly* concerns only aspects of non-mental reality. The rule will be required to facilitate transitions from knowledge of those aspects to conclusions about one's mind. So if the accomplishment of the former knowledge essentially requires certain kinds of collateral self-knowledge of mental states and properties, the Byrnean account offered will carry the kind of reductive epistemic payload which transparency aims at only if any self-knowledge essentially involved in achieving knowledge of the relevant antecedent, is independently accounted for.

Say that a Byrnean rule is *illicitly demanding* if it doesn't meet this condition— if a necessary condition for the achievement of knowledge of its antecedent, even if no mention of anything overtly mental is contained therein, is that the agent has, or gets, certain kinds of psychological self-knowledge. Then if these include the very sub-species of self-knowledge supposedly being accounted for, the Byrnean account is tacitly circular. And if they include only other kinds of self-knowledge, then a properly reductive account of those has to be given before any claim can be sustained to have delivered a theory which does full justice to the intuitive idea that the world within is available to knowledge that draws only on an 'outward gaze'. The objection, then, is that BEL, as 'non-perversely' understood, is illicitly demanding.

It is important to see that this objection would not be addressed if instead of a conditional rule, it were proposed that the gist of BEL is better captured by something on the model of a natural deductive rule of proof: roughly,

> If you have arrived at a line, P, at which all assumptions have been discharged, you may infer the necessitation of P.

Correspondingly, for BEL, Byrne could have proposed something like: the transition is acceptable from recognition that P to the claim that you believe that P.[17] Well, to be sure: on this proposal, there would no longer be the worry that afflicted the conditional formulation of BEL, about our insouciant attitude to unheeded truths. But the evident continuing problem would be that in order to follow this new version of BEL in any particular case, you will need to keep track of your epistemic situation and achievements — everything involved in knowing whether you have recognized that P. Keeping track will once again implicate a suitcase of attitudinal self-knowledge, including knowledge of a range of other beliefs.

All that said, it would be inappropriate to conclude on anything but a note of admiration for Byrne's book, which is full of ingenious moves and proposals, interesting arguments and is written throughout with good humor and gusto.

---

[17] In "Introspection" (2005), Byrne is explicit about the epistemic rule BEL functioning like a *proof rule* rather than a rule of inference (see pp. 95-96). Indeed he also suggests the same on pp. 103-104 of *Transparency and Self-Knowledge*, and in particular footnote 7 on page 104.

The exploration and criticism of other contemporary views in the early chapters will be especially valuable for anyone wishing to teach or introduce themselves to the contemporary debates about self-knowledge. Moreover, when so much recent work on the philosophical problems posed by self-knowledge has preceded in a somewhat methodologically unselfconscious fashion, it is refreshing and helpful to encounter a systematic treatment which is rigorously disciplined in the fashion illustrated by Byrne's deployment of his four overarching constraints— even if they may be controversial in detail. It merits acknowledgment, further, that the character of one's self-knowledge of one's attitudes — knowledge that in basic cases seems no less authoritative and immediate than basic self-knowledge in general yet is characteristically uninformed by any distinctive phenomenology — can seem especially mysterious. The great attraction of Evans' ur-thought is its promise to dispel some of the mystery about this phenomenological 'blankness' or apparent baselessness. Byrne's development of Evans' idea is, in our view, the most interesting and striking to date. But for all the considerable dialectical resourcefulness manifested in the details of his treatment, we do not, for the reasons outlined or prefigured in the preceding, think his project is or can be successful, even for the masthead case.[18]

University of Stirling and New York University.

[18] The research for this study was conducted as part of the work for the project, *Knowledge Beyond Natural Science*, which was funded by the John Templeton Foundation at the University of Stirling from March 2017 to November 2019. We gratefully acknowledge the support of the Templeton Foundation and the valuable comments and criticisms of our colleagues in the project, including , Philip Ebert, Adrian Haddock, Jonathan Jenkins Ichikawa, Carrie Ichikawa Jenkins, Indrek Lobus, Alisa Mandrigin, Giacomo Melis, Alan Millar, Sonia Roca Royes, Sam Symons, Peter Sullivan, Joshua Thorpe , Xintong Wei, Mike Wheeler and the invaluable input and advice from the project's two Academic Auditors, Jim Pryor and Alex Byrne himself.

Appendix: Byrnean Rules for mental states and characteristics other than Belief

To extend his account of self-knowledge beyond the case of belief, Byrne invokes a variety of epistemic rules that share with BEL the general form:

R If conditions C obtain, believe that P,

where P attributes some mental state to the subject. Our aim in this appendix is to provide a short synopsis of these rules and prefigure some of the problems we believe they raise. Some of the worries we canvassed above for BEL also, in our judgement, apply equally to the additional rules. For example, at least some of the rules require that the subject form beliefs based on 'terrible' reasons. And at least some of the rules are ones that we can't be plausibly described as following. However, we will not press these worries further here.

Let us begin with the case of visual perception. Byrne focuses on self-knowledge that one is seeing an F – for example, self-knowledge that one is seeing a hawk. Obviously, it would be hopeless to suggest that one acquires such knowledge via a rule like:

HAWK If there is a hawk over there, believe that you see a hawk

Given Byrne's understanding of what it is follow a rule (TSK, pp. 101-102), one follows HAWK on a particular occasion if and only if, on that occasion, one believes that one sees a hawk because one recognizes that there is a hawk over there. But a belief formed on such a basis would only sometimes be true: there are countless possible cases where one might recognize that there is a hawk over there without *seeing* any hawk (one might e.g. see a hawk's nest and hear fluttering nearby or simply be told that there is a hawk over there).

Byrne's fix to this problem involves strengthening the antecedent of HAWK. Call a *v-proposition* any proposition that can constitute the content of a visual experience. Byrne suggests that the right rule will involve transition from recognizing the truth of a v-proposition concerning a thing that one takes to be a hawk to the belief that one sees a hawk. More generally, self-knowledge that one is seeing an F is to be explained in terms of the following of the schematic rule:

SEE If $[\ldots x \ldots]_V$ and x is an F, believe that you see an F

We are instructed to think of the v-proposition that $[\ldots x \ldots]_V$ as describing a certain state of affairs "in the language of vision" (TSK, p. 140). Byrne's idea is that because the language of vision is unwritten and unspoken, the only way in which one can bring oneself to recognize that a v-proposition about x is true is by having a veridical visual experience involving x. Thus, in all ordinary situations, following SEE will result in the formation of a true belief.

This proposal relies on several substantial assumptions. First, it assumes that visual experiences have conceptual content in the first place – the kind of content that one can believe as part of the process of drawing an inference – and that the kind of conceptual content visual experiences have is indicative of its provenance: whenever one believes such a content, one must be having a visual experience. Call these assumptions *conceptualness* and *indicativeness*,

respectively. Second, since following SEE requires recognizing (hence, believing) that the condition specified in the antecedent of the rule is satisfied, following SEE requires believing the content of one's visual experience. To rule out possible counterexamples to this requirement (most notably, cases of known illusion), Byrne endorses *belief-dependence*, the view that vision constitutively involves belief in the relevant v-proposition (TSK, p. 144). Finally, Byrne admits that, unlike BEL, SEE is only practically self-verifying, meaning that it generates true beliefs, not in all situations, but in all *ordinary* situations (TSK, p. 140). This commits him to the view that our epistemic access to our own perceptions is less secure than our epistemic access to our own beliefs. Call this view *disparity*.

Each of conceptualness, indicativeness, belief-dependence and disparity may be plausibly called into question. But even granting all these assumptions, a potential problem arises concerning the explanatoriness of Byrne's proposal.

Consider a belief generated by SEE. That belief wouldn't be safe – hence, on Byrne's account, it wouldn't amount to knowledge – if the subject could easily have formed it falsely. In particular, the belief wouldn't be safe if the subject could easily have formed it in a way based on recognizing the truth of some *non-visual* proposition – for, in that case, there would be no guarantee that the subject is seeing anything at all. This means that Byrne's account can only explain self-knowledge of perception on the assumption that the subject who follows SEE is reasonably good at telling the difference between its being the case that [. . . x . . .]$_V$ and its being the case that [. . . x . . .], where the latter proposition describes the same state of affairs as the former (or a state of affairs as similar to the former as is allowed by the difference between the two propositions) but *not* in the language of vision.

The problem is that it's hard to see how this ability – telling visual facts and propositions from their non-visual counterparts – can be justifiably described as the kind of outward-looking ability that one can expect to be invoked (and taken for granted) in the context of a 'transparency' account of self-knowledge. It's true that the v-proposition that [. . . x . . .]$_V$ concerns the outer world, not the subject. But how is one supposed to tell whether *that* proposition is true, rather than its non-visual analogue, if not by looking inward at whether it's a proposition that constitutes the content of a visual experience one is having?

Byrne's account of self-knowledge of sensation, imagination, memory and occurrent thought has a similar shape – and is open to the same kind of objection.

In the case of sensations of pain, Byrne's account invokes *p-propositions*, propositions "concerning qualities of painful disturbances occurring in the bodies of animals" (TSK, p. 149). The suggestion is that, when one feels a pain, one is perceiving a bodily disturbance via a special kind of perception called 'nociception'.[19] Since, in creatures like us, nociception happens to be "an exclusive conduit" (TSK, p. 150) for facts expressed by true p-propositions, believing one of these propositions is a guarantee that one is nocicepting, i.e. feeling a pain. This makes the following rule practically self-verifying:

[19] This has the implication that there can be illusions and hallucinations of pain. Byrne happily accepts this implication (TSK, pp. 154-5).

(PAIN) If [. . . x . . .]$_P$, believe that you feel a pain.

Similarly, to explain self-knowledge of visual imaginings, Byrne assumes that "the content of visualizing is the same kind as the content of vision – albeit degraded and transformed in various ways" (TSK, p.188). Letting '« . . . x . . . »$_V$' schematically express a *v-proposition*, i.e. a degradation and transformation of a v-proposition, he suggests that one comes to know that one is imagining a duck via:

(IMAG-DUCK) If « . . . x . . . »$_V$ and that$_x$ is a duck, believe that you are imagining a duck

Assuming further that (visual) episodic memory constitutively involves visualizing, Byrne explains self-knowledge of visual episodic memory in terms of epistemic rules like:

(MEM-DUCK) If « . . . x . . . »$_V$ and a duck was that^x way, believe that you are recollecting a duck

Finally, Byrne explains self-knowledge of occurrent thought about by invoking *s-propositions*, the propositions of inner speech (which he suggests we can think of as degraded and transformed versions of the propositions of outer speech). Letting « . . . . . »^e_S schematically express some such proposition, the suggestion is that one comes to know that one is thinking about a certain object o by following:

(THINK) If « . . . . . »^e_S and that$_e$ is about o, believe that you are thinking about o

In each of the cases just mentioned, Byrne is committed to analogues of conceptualness, indicativeness, belief-dependence and disparity. Conceptualness is especially controversial in the case of pain – a mental state that, according to some, is not appropriately described as having content, whether conceptual or not. Belief-dependence is especially controversial in the case of memory and imagination: since, in believing the antecedents of rules like IMAG-DUCK and MEM-DUCK, one must believe the relevant *v-propositions*, and such propositions are taken to concern 'shadowy, insubstantial, ghostly' objects (TSK, p. 193), the claim that we follow those rules commits Byrne to saying that "people generally harbor harmless delusions about a shadowy world of images'" (ibid.). Finally, exclusivity is especially controversial in the case of inner speech – for one might have thought that the propositions which constitute the content of inner speech are the same as the propositions that constitute the content of belief.[20]

In addition, it is not obvious that the rules above can guarantee a robust form of privileged access. Take MEM-DUCK. Even in fairly ordinary situations,

[20] Since employing the demonstratives that occur in the antecedents of IMAG-DUCK, MEM-DUCK and THINK requires attention to what one is visualizing or uttering in inner speech, there's also a worry that these rules might be illicitly demanding, in the sense defined in § IV.

one may recognize the truth of the antecedent of MEM-DUCK without recollecting any duck: for example, I may visualize my uncle's duck cleaning its feathers a certain way and know (merely based on the fact that it's statistically probable) that last month my uncle's duck did clean its feathers that way, without it being true that I am recollecting my uncle's duck. If I formed the belief that I am recollecting a duck by following MEM-DUCK, I would be liable to go wrong in many quite ordinary kinds of circumstance.

Byrne's account of emotion, desire and intention has a different shape – in these cases, the proposed inference isn't 'mad' and the antecedent of the rule doesn't involve *sui generis* propositions. Instead, the rules suggested are the following:

> INT If you will ϕ, believe you intend to ϕ
>
> DES If ϕing is a desirable option, believe that you want to ϕ
>
> DIS If x is disgusting, and produces disgust reactions in you, believe you feel disgust at x

We cannot hope to do justice here to Byrne's engaging discussion of these rules. But it is instructive to consider some of the salient difficulties they face.

INT generates false beliefs in all cases where one recognizes that one will ϕ without intending to ϕ: for example, I may recognize that I will die, without intending to die. To address this problem, Byrne appeals to Anscombe's idea that self-knowledge of one's intention is arrived at 'without observation'. Glossing 'knowledge without observation' as 'knowledge not resting on evidence' he suggests that "one will not follow INT if one believes that one's belief that one will ϕ rests on good evidence that one will ϕ'" (p. 171). But this means that, in order to decide whether to follow INT, one must know on what basis one holds the belief that one will ϕ – a kind of self-knowledge of which Byrne gives us no account in the book.[21] The objection of illicit demandingness arises again.

DES faces the familiar objection that we may desire things that are not desirable and fail to desire things that are desirable (cf. Cassam 2014, chapter 1). Byrne responds to this objection by claiming that "the relevant sense of 'desirable' is easy to miss" (p. 166) and that DES can be defeated (specifically, "one will not follow DES and conclude that one wants to ϕ, if one believes that (a) one intends to ψ, (b) that ψing is incompatible with ϕing, and (c) that ψing is neither desirable nor all-things considered better than ϕing" (165)). These adjustments may succeed in blocking the objection, but they also make Byrne's proposal less appealing than it might seem at first. If the relevant sense of 'desirable' is easy to miss, the question whether ϕing is a desirable option is easy to misunderstand. But the question whether I want to ϕ is not easy to misunderstand. Why suppose that we settle a question that is not easy to misunderstand in terms of a question that is?

---

[21] Byrne assumes the Williamsonian thesis that one's evidence is one's knowledge (TSK, p. 2) and that one can know that one knows that p by following the rule 'If p, believe that you know that p' (TSK, p. 116). But these assumptions explain, at most, how one knows that the proposition that p forms part of one's evidence. They don't explain how one knows that a certain belief is (not) *based* on that evidence, when it is (not).

DIS has a distinctive Rylean flavor. The suggestion is that finding out whether one feels disgust at something involves attending to sensational and behavioural cues produced by that thing (e.g. feelings of queasiness and distinctive facial expressions). But one obvious problem is that the disgust reactions mentioned in the antecedent of DIS might be produced by x in the wrong way. For example, suppose I am told that a disgusting object contained in a sealed box is producing disgust reactions in me via some electrodes placed in my brain. In such a situation, I may recognize that the object is disgusting and is producing disgust reactions in me – yet it would be false to say that I feel disgust *at* the object. Of course, the point is not that the situation just described is an ordinary one – it isn't. The point is that, presumably, even if in the situation just described, I would retain privileged access to whether I feel disgust *at* the object in the box. Consequently, that access cannot be satisfactorily explained by DIS.

University of Stirling and New York University

Bibliography

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, 1st edn. London: Routledge & Kegan Paul.

Bar-On, D. (2004) *Speaking My Mind: Expression and Self-Knowledge*, 1st edn Oxford: Oxford University Press.

Bilgrami, A. (2012) *Self-Knowledge and Resentment*, 1st edn. Cambridge: Harvard University Press.

Boghossian, P. (2014) What is inference?, *Philosophical Studies* 169, pp. 1-18

Boyle, M. (2009) Transparent self-knowledge. *Aristotelian Society Supplementary Volume* 85: 233–41.

Byrne, A. (2005) Introspection, *Philosophical Topics*, 33(1), pp. 79–104.

Cassam, Q. (2014) *Self-Knowledge For Humans*, 1st edn, New York: Oxford University Press

Coliva, A. (2016) *The Varieties of Self-Knowledge,* 1st edn, London: Palgrave Macmillan

Coliva, A. (2017) *How to be a Pluralist about Self-Knowledge*, pp. 253-284 in *Epistemic Pluralism*. 1st Edition, Edited by A. Coliva and N. Jang Lee Linding Pedersen.

Davidson, D. (1984a) First person authority. *Dialectica* 38, pp. 101–11.

Evans, G. (1982) *The Varieties of Reference*. 1st edn. Edited by J. H. McDowell. New York: Oxford University Press.

Fernandez, J. (2013) *Transparent Minds: A Study in Self-Knowledge*. 1st edn. Oxford: Oxford University Press

Harman, G. (1986) *Change in View*. Cambridge, MA: MIT Press..

Moran, R. (2001) *Authority and Estrangement*. Princeton, NJ: Princeton University Press.

Nichols, S., and S. Stich. (2003) *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.

Squire, L. R. 1992. 'Declarative and nondeclarative memory: multiple brain systems supporting learning and memory'. *Journal of Cognitive Neuroscience* 4: 232–43.

Wright, C. (1987). 'On making up one's mind: Wittgenstein on Intentions'. In P. Weingartner and G. Schulz (Eds.), *Logic, Philosophy of Science and Epistemology, Proceedings of the XIth International Wittgenstein Symposium,* Vienna, Holder-Pickler-Tempsky, pp. 391-404.