Byrne on transparent introspection

Michael Roche, Department of English and Philosophy, Idaho State University, USA, rochmich@isu.edu

KEYWORDS: Alex Byrne; transparency; self-knowledge; introspection; introspective uniformity; introspective dissociations

Transparency and Self-Knowledge, by Alex Byrne, Oxford University Press, Oxford, UK, 2018, xi+227 pp., £46.61 (Hardback), ISBN 9780198821618

Some philosophers have what we might call the "transparency intuition". This is the sense that we introspectively know our mental states by attending outwardly, towards the world. Gareth Evans' (1982) example is perhaps familiar. By considering whether there will be a third world war, I can know that I believe there will be a third world war. No inner glance is needed. Evans also suggests, oversimplifying a bit, that by considering how things are at this place now, I can know, e.g., that I am perceiving a partially filled coffee mug.

Yet this world-to-mind procedure seems epistemically problematic. Understood as an inference, it can seem "mad" (Boyle 2011), for the worldly premise provides very weak support for its psychological conclusion. Propositions about wars and mugs are evidentially disconnected from those about one's psychology. The connection can be strengthened by supplementing the worldly premise with facts about oneself, but the transparency intuition does not depend on this.

This is the *puzzle of transparency*. It consists of the transparency intuition plus the epistemic concerns just noted. Alex Byrne aims to solve this puzzle in his excellent book *Transparency and Self-Knowledge*. The book's first four chapters set the stage by introducing the topic of self-knowledge, explaining the transparency intuition, critically examining extant theories of introspection, and surveying previous discussions of the puzzle. Byrne's solution begins in Chapter 5, where he discusses belief introspection. It is then extended to other mental kinds in the three remaining chapters.

I

Byrne defends a "straight solution" (p 77) to the puzzle whereby the transparency intuition is vindicated and the epistemic concerns are overcome. World-to-mind transparency inferences, Byrne argues, produce introspective knowledge. Introspective knowledge is self-knowledge that is both "privileged" in that it is epistemically more secure than knowledge of others' mental states (p 5) and "peculiar" in that it is obtained in a way not available to others (p 8).

A noteworthy feature of Byrne's solution is its generality. He argues that *all* introspective knowledge is transparent. I return to this generality later on. First, however, it will be helpful to consider his discussion of *belief* introspection. This constitutes the centerpiece of his theory.

Byrne's proposal, put simply and omitting jargon, is that introspective knowledge of the form, *I believe that p*, is inferred from a single premise of the form, *p*. This is meant to capture the transparency intuition, given that the lone premise will typically concern the world (e.g., *there will be a third world war*).

Inferring a conclusion from a premise entails believing that premise, Byrne claims. Consequently, when I self-ascribe a belief by the transparency inference, that very belief is part of the inferential process. The inference is thus very safe. It can go wrong only in the unlikely

event that belief in the premise is lost before the inference is completed (p 104). (Such a change of mind is unlikely given that the transparency inference has just a single step.) Because the aforementioned connection holds regardless of whether the (believed) premise is true, the inference is no less safe when the premise is false. If I believed the Earth is flat, I could use the transparency inference to know this.

In virtue of the inference's safety, Byrne alleges that it produces privileged self-knowledge (pp 109-112). Additionally, because *my* reasoning from a premise does not entail that *you* believe that premise, the inference cannot be used to know others' beliefs. The inference is thereby alleged to produce peculiar self-knowledge as well (pp 108-109).

Two challenges arise when attempting to extend this theory to other mental kinds. To illustrate, consider *desire*. First, there is no plausibility to the idea that I infer that *I desire that p* from the premise that *p*. Knowing that one has a desire without believing that desire's content is commonplace. For example, I might know that I want a beer while believing that I do not have one (p 156). Second, drawing an inference from a premise does not entail that one has a desire related to that premise's content (p 161). In these ways, the transparency inference for desire will be unlike that for belief. To show that it is especially safe, a different rationale will be needed.

Byrne proposes that we infer our desires from premises about what is *desirable* (p 161), understood "… in the *Oxford English Dictionary* sense of having 'the qualities which cause a thing to be desired: Pleasant, delectable, choice, excellent, goodly'" (p 160). To judge that an action or state of affairs is desirable in this sense is to make a judgment about the world, not one's mind. Byrne takes this inference to be safe since "… one's desires tend to line up with one's *beliefs* about the desirability of the options …" (p 162, original emphasis). Whether a mere

tendency for alignment is strong enough to accommodate *privileged* self-knowledge is unclear, however. For as misalignment increases, so too does the opportunity for error.

Byrne's treatment of *seeing*—which precedes his discussion of desire—is in some ways more convincing. He takes visual experiences to have propositional contents. These contents, which he calls "v-propositions", are characteristically rich and concern "… shape, orientation, depth, color, shading, texture, movement, and so forth …" (p 136). A v-proposition cannot, as a practical matter, be expressed linguistically but can be represented as "[…x…]$_V$", where "x" stands for an object (e.g., a hawk). Introspective knowledge of the form, *I see an F*, Byrne claims, is inferred from the conjunctive premise: *[…x…]$_V$ and x is an F* (p 139).

The richness of v-propositions guarantees, practically speaking, that I can reason from a v-proposition only when having a visual experience with that v-proposition as its content. Apparent memories of v-propositions are too degraded to truly have v-propositions as their contents. Consequently, I do not use the transparency inference when reasoning from memory (p 141). The transparency inference for seeing thus appears to be very safe, and it is alleged to produce privileged knowledge as a result. Because *my* reasoning from a v-proposition does not practically guarantee that *you* are having a visual experience with that content, the inference is also alleged to produce peculiar knowledge.

Importantly, the capacity to use transparency inferences, such as those already described, is alleged to be a byproduct of normal human rationality (p 112). Psychological mechanisms specialized for acquiring self-knowledge are not needed. Byrne's theory is thus "economical" in that it tries to "… explain self-knowledge solely in terms of epistemic capacities and abilities that are needed for knowledge of other subject matters" (p 14). Inner-sense theories, by contrast, are

neither economical nor inferential. Some such theories posit perception-like mechanisms for monitoring the mind. Economy, which I return to later, is an attractive feature of Byrne's theory.

Some of the most interesting parts of Byrne's book are his attempts to extend his theory to non-beliefs. In addition to desires and visual states, he extends his theory to pains, intentions, disgust, memories, imaginings, and thoughts. (He uses seeing, pain, and disgust as examples of perception, sensation, and emotion.) This is a bold undertaking, to say the least.

These extensions are not without their problems, however. I noted above that whether the desire inference can produce privileged knowledge is unclear. Additionally, the extensions often lead to surprising and controversial consequences. Two such consequences arise when extending the theory to seeing and pain. Byrne ultimately claims that seeing implies believing (pp 142-146) and that phantom pains are not pains (pp 151-155). Still, the discussions are always fascinating and enlightening.

I suggested above that the extension of the transparency approach to seeing is more successful than that to desire. This is somewhat unsurprising, for if the transparency intuition is strongest for belief, then perception is a close second. It is thus no coincidence that Byrne's first two extensions are to perception and sensation—where the latter is understood on a perceptual model. Could it be that transparent introspection applies *only* to these mental kinds (plus belief)? Byrne is aware of this concern. After completing the first two extensions, he writes: "[w]e have now reached the book's Rubicon. Is this where transparency runs out …?" (p 155). He answers negatively. Next, I scrutinize the reasoning behind this answer.

II

Byrne's crossing of the Rubicon is aided by a short but important argument from the start of Chapter 7. The argument concludes that introspection is *uniform*, that is, all introspective

knowledge is explained in the same basic way (p 157). This implies that if beliefs and perceptions are introspected transparently, then so too are all other introspectable mental kinds. The argument is thus important in motivating Byrne's later attempts to extend his theory.

The argument can be stated simply (pp 157-158): (1) if introspection is not uniform, then we should expect *introspective dissociations*; these are cases where a person has introspective access to some but not all introspectable mental kinds; (2) introspective dissociations do not occur; thus, (3) introspection *is* uniform. The reasoning behind premise two is quite simple: "… such conditions do not seem to occur" (p 157). The reasoning behind premise one is more complex.

Consider a particular way in which introspection might be *dis*unified. Suppose, for example, that beliefs and perceptions are introspected transparently (à la Byrne), but desires, intentions, emotions, memories, imaginings, and thoughts are introspected non-transparently, say, by inner sense. Suppose, further, that there is no overlap; beliefs and perceptions can *only* be introspected transparently, while desires and the rest can *only* be introspected by inner sense.

Such disunity seems to allow for selective damage. Neurological and/or psychological impairment could disrupt one but not both types of introspection, thereby undermining introspective access to one set of mental kinds but not the other. This would be analogous to losing access to visual, but not auditory, information due to damage to one's vision, but not hearing (pp 157-158). This is the reasoning behind premise one.

Unfortunately, the argument fails. Both its logic and premises are questionable. I shall restrict my attention to premise two, however. My complaint can be stated simply: we cannot be confident that introspective dissociations do not occur. For if they did, they would not be easily detected. Consequently, we cannot be confident that premise two of Byrne's argument is true.

The difficulty of detecting introspective dissociations can be seen by returning to the above example of disunity. There are two types of dissociation to consider, one caused by disruption to inner sense and the other caused by disruption to transparent introspection. I begin with a case of the first.

Imagine a person whose inner sense becomes completely inoperable but whose transparent introspection is spared. Although she can introspect her beliefs and perceptions, she is unable to introspect all other mental kinds. Whether this would be detected is not at all obvious. For she could still self-ascribe mental states using her ways of ascribing them to others. She could thus reasonably, and even knowingly, self-ascribe mental sates belonging to those mental kinds to which she lacks introspective access. Although these self-ascriptions would be third-personal—relying on behavioral and circumstantial evidence about herself—this need not be obvious to her. After all, confabulations are third-personal but can feel first-personal.

More importantly, she could reasonably (even if not knowingly) self-ascribe mental states *transparently*, as Byrne describes. That we have the capacity to reason transparently is not in doubt; opponents of transparent introspection merely deny that such reasoning produces introspective knowledge. Consequently, she will not be limited to self-ascribing mental states third-personally. Given these resources, there is reason to doubt that her dissociation would be detected, either by herself or by others.

Next, consider a case of the second type of dissociation. Imagine a person whose transparent introspection becomes completely inoperable but whose inner sense is spared. She is unable to introspect beliefs and perceptions but can introspect all other introspectable mental kinds. This type of dissociation is importantly different than the first. Because transparent introspection is inferential and economical, this type of dissociation implicates more than just introspection. It

would apparently stem from general impairments of rationality, reason, and perhaps communication. That these impairments would overshadow the resulting introspective disruptions, making them difficult to detect, is entirely possible.

The issues here are complicated, of course, and a full defense of my criticism requires more detail. Still, I think it is clear that the argument's second premise needs more defense. If we cannot confidently rule out the occurrence of introspective dissociations, then we cannot confidently rule out the possibility that transparency *does* run out at belief and perception.

Without this argument, readers skeptical that transparency applies across the board will have little reason to follow Byrne as he extends his theory. For such readers, the book's final chapters may feel too hypothetical: *if* all introspection is transparent, then this is how it could work. Relevant here is the fact that Byrne takes himself to be describing introspection as it is, not prescribing how it could be. This leads to a final worry.

Missing from Byrne's theory is an explanation of *why* we engage in transparent reasoning. He is clear that such reasoning typically occurs unconsciously (p 114, p 124), and that self-knowledge is useful (p 114). This suggests that we are disposed to unconsciously engage in transparent reasoning due to its utility. But this leaves unexplained the disposition's origin. Is it innate or acquired? Nor does it explain the conditions under which it is exercised.

At stake here is the theory's economy. If we are innately disposed to unconsciously self-ascribe mental states in response to judgments about the world, then the gap between Byrne's theory and its non-economical rivals seemingly narrows. If the disposition is instead claimed to be acquired, then an explanation of its acquisition is needed, especially considering the oddness—or even "madness"—of transparency inferences.

III

Despite these worries, I highly recommend Byrne's book. It is clear, engaging, instructive, creative, ambitious, and well-argued. Byrne's work on transparent introspection has already been extremely influential in the self-knowledge literature. This book will only solidify that influence.

Boyle, M. (2011), "Transparent Self-Knowledge", *Aristotelian Society Supplementary Volume* 85, 233-41.

Evans, G. (1982), *The Varieties of Reference*, J. McDowell (ed.), Oxford: Oxford University Press.