Why Transparency Undermines Economy

This is the penultimate draft of a paper forthcoming in *Synthese*.

Abstract

Alex Byrne (2005; 2011a; 2011b; 2011c and 2012) offers a novel interpretation of the idea that the mind is transparent to its possessor, and that one knows one's own mind by looking out at the world. This paper argues that his (2011c and 2012) attempts to extend this picture of self-knowledge force him to sacrifice the theoretical parsimony he presents as the primary virtue of his account. The paper concludes by discussing two general problems transparency accounts of self-knowledge must address.[1]

1. Introduction

How do we know the contents of our own minds? One answer is that our own minds are transparent to us; we know simply by inquiring about the world (Evans 1982). I can treat the question 'Do I believe that *p*?' as equivalent to the question *whether p*, and by treating the two questions as equivalent, I am almost guaranteed to answer to the first question correctly.

Alexander Byrne (2005 and 2011a) has argued for a particular elaboration of transparency, according to which we know our own minds through an inference rule. All of us (or at least all normal humans) reason according to the following principle:

BEL: If *p,* then believe that you believe that *p*!

Now this principle may seem dubious—I do not believe everything that is the case—but Byrne points out that using it is guaranteed to produce true conclusions anyway. My premises must be the propositions I believe. Byrne thus calls the principle "strongly self-verifying." Because the principle is strongly self-verifying, conclusions reached on its basis are justified; they count as knowledge.

The problem is that transparency looks most plausible when we focus on one particular attitude, belief. A person's beliefs are her picture of what the world is like, and so it is reasonable that we could know them by asking ourselves what the world is like. But optical illusions can persist despite our better knowledge, and thus we cannot know how these illusions appear to us by asking ourselves how the world is—since the world visually appears to us other than the way that, as far as we are concerned, the world is. Intentions, for their part, seem to have the role of changing the world, rather than representing the way the world is (cf. Anscombe 1957; and Humberstone 1992)

In his (2011c) Byrne offers an extension of this inferentialist account to knowledge of one's own intentions; and in his (2012), Byrne offers a similar extension for knowledge of one's own sensory experiences. If his arguments are successful, then his inferentialist version of transparency begins to look extremely credible as a general explanation of the possibility of self-knowledge. The primary virtue of his account is its theoretical simplicity. As he puts it:

> …[T]he transparency account is *economical*: it explains self-knowledge in terms of epistemic capacities and abilities that are needed for other subject matters. (2012: 207)

In other words, this account explains how people can know their own mental states without positing any special capacities of self-knowledge (as in Lycan 1996), and without making self-knowledge radically discontinuous with other forms of knowledge (as in McGeer 1996; Moran 2001; Boyle 2011; and arguably Valaris 2014). At the same time, it does not make knowledge of one's mental states essentially third-personal, so that one infers what one thinks on the basis of one's behavior (cf. Carruthers 2013). It preserves, on the one hand, the idea that one's epistemic access to one's own mental states is privileged; at the same time, all that people need for this privileged knowledge is capacities for knowing the external world, plus inference rules.

Byrne's account has been criticized on the grounds that those supposedly utilizing the inference rules would regard them as fallacious—anyone with enough sense to recognize her own fallibility does not think that it follows from the truth of p that she believes p—and this casts doubt on whether a person disposed to believe that she believes that p whenever she believes that p is engaged in a process rational enough to be regarded as genuine inference (cf. Boyle 2011; Moran 2012; and Shoemaker 2012; also see Valaris 2011). More recently, Lauren Ashwell (2013) has argued that Byrne's (2011b) account fails as an explanation of knowledge of one's own desires—and only succeeds if we supplement his theory with commitments that undermine its claim to theoretical economy.

This paper continues the second line of argument with respect to his account of intention; in the case of perception, on the other hand, it argues that Byrne has already

taken on commitments incompatible with the goal of economy. Thus, while his attempts to extend the account to perception and intention are ingenious, they rob the account of any claim it might have to theoretical economy, the virtue that was supposed to draw us to inferentialism in the first place. These two objections will illustrate challenges for transparency-based accounts of self-knowledge more generally.

2. Knowledge of Perception

How does Byrne's account apply to perception? We start with the assumption that the content of perception is propositional. I see *that a chair is a few feet in front of me*; I smell *that coffee is brewing*. Byrne makes a further claim about the content of perception: each perceptual mode takes for its content a special class of proposition, one appropriate to the modality. So the content of visual perception is a visual proposition (or *v-proposition*), whereas smell takes for its content olfactory propositions (or *o-propositions*), touch takes tactile propositions, and so on (2012: 197-201).

Byrne only provides a detailed account of vision and its relation to v-propositions, though presumably the relation other senses bear to their respective contents will be analogous. For Byrne, propositions in general can be thought of as ways that the world can be. V-propositions, then, are ways that the world can be *with respect to* its visible objects and properties, whereas o-propositions are ways that the world can be with respect to objects and properties detectable through smell. It is important to emphasize that the propositions in both cases are not specifically visual or olfactory *representations* of how the world can be; rather, they are what visual or olfactory experience represents—they are the ways the world can be, specified with respect to those properties accessible to the relevant modality. Byrne is explicit that vision, for example,

represents "v-facts" (2012: 199-200).[2] The theoretical importance of this commitment will become clear momentarily; for now it is enough to note that this is a standard commitment about the nature of propositions.

The other distinguishing characteristic of v-propositions is complexity. A visual perception of a green leaf, or a silver spoon, or a yellow taxicab presents us with vastly more information about the hue, shade, and spatial relations among the visible parts of the object than we can express in language. A picture is worth a thousand words, and then some:

> …[G]iving a theoretically satisfying characterization of v-facts is difficult. …Complexity or informational richness is no doubt part of the story, but even in the case of viewing a very simple scene—say, a red spot against a grey background—it is unclear just how to proceed. Just concentrating on one feature of the spot, its hue, the predicate 'is red' (or even some made-up predicate like 'is red$_{29}$') does not quite do it justice. The particular red hue of the spot might be a little yellowish, or alternately a little bluish… (2012: 199-200)

If we think of propositions as sets of possible worlds, we can treat v-propositions as sets formed by partitioning possible worlds solely on the basis of visible differences. These visibility-based partitions must be more fine-grained, however, than the partitions that can be expressed in natural language, or than can be made in ordinary belief (visual beliefs will be dealt with in a moment).

---

[2] "Vision, we may say, reveals the *visual world:* the world of v-facts. In the visual world things are colored, illuminated, moving, and so on, but not smelly or noisy" (2012: 200). Byrne goes on to add: "Vision is, at least in creatures like us, an exclusive conduit for v-facts" (ibid.).

While Byrne does not pursue this option, it's worth noting his v-propositions could be given a Fregean gloss as well (though a commitment to Fregeanism may sit poorly with his ultimate aim of theoretical economy).[3] In this case, the distinguishing feature of such propositions is not that they are ways the world could be *with respect to visible properties,* but that they are ways the world could be *under a visual mode of presentation* (Chalmers 2006). On the former understanding of v-propositions, the visual experience that *the taxicab is yellow* and the belief that *the taxicab is yellow* have different contents, because the visual experience takes for its content a subset of the set of possibilities that make up the belief. On the latter understanding, the contents differ in roughly the same way that the contents of that belief that *x is Hesperus* and the belief that *x is Phosphorus* differ: the contents are composed of different *senses* (or something very similar to senses)*,* even if they have a common referent.

Either interpretation of v-propositions raises questions, though, about the relation of my seeing that the taxicab is yellow with my ordinary, assertible belief that the taxicab is yellow. While Byrne does not spell out the exact relation of v-propositions to other propositions, we can extrapolate the following from his above commitments along with other arguments he goes on to make. I will use p_V to name visual propositions, whereas non-visual propositions will simply be identified as *p*. The v-proposition *the-taxicab-is-yellow$_V$* entails *the-taxicab-is-yellow.* The first proposition will pick out some much more specific way that the taxicab is yellow and will presumably include additional visible properties of the taxi, such as the shape of the chassis. Thus, if that proposition is true, the non-visual proposition must be true as well.[4] There is still a question of how the

---

[3] Thanks to an anonymous referee for suggesting this, as well as noting the problem it might raise for the theory's parsimony.

[4] These points will apply, *mutatis mutandis,* to the Fregean interpretation of v-propositions.

perception with the first proposition as content justifies one's belief in the second—but a number of answers are consistent with the picture sketched out here.

With the v-proposition, Byrne can construct inference rules by means of which we know our visual experiences (2012: 199):

SEE: If p_V then believe that you see that p_V!

In other words, one takes the world to be some visibly accessible way of sufficient visible complexity, and concludes on that basis that one sees the world to be such a visibly accessible, complex way. Similar rules could be offered for the other sense modalities.

With SEE in place, we can see why Byrne is committed to his various claims about v-propositions. Something must distinguish them from non-v-propositions, otherwise I will apply SEE to every proposition I believe, and so conclude that I see that Henry VIII had six wives. At the same time, Byrne's inferentialist account of self-knowledge is supposed to be a more precise way of capturing the idea that the mind is transparent, and that one knows one's mind by looking out at the world. This rules out identifying v-propositions with specifically visual ways of encoding information, or specifically visual ways of representing the world. Imagine this were not the case. Then, for any given proposition p which I held true, to know whether I should conclude that I see that p on its basis—that is, to know whether it could serve as a premise in SEE—I would first have to identify how I was representing p. But this would mean that application of the inference rule required prior identification of what sort of representational states I was in. Byrne thus identifies v-propositions a particular subclass, not of ways the world could be represented, but simply of the ways that it could be—in particular, ways that its visible properties and objects could be.

But I may have been blinded, yet believe that there is a brown cow standing in front of me thanks to a friend's report. If SEE told me to infer that I see a brown cow in such a situation, it would not be strongly self-verifying. Byrne's appeal to the complexity of visual propositions is meant to prevent such inferences. Restricted to these propositions, SEE, is a *practically* self-verifying principle: if an introspecting subject had access to "the—as-yet-unwritten—language of vision" application of SEE would result in false conclusions; but in all ordinary situations these inferences will generate truth (2012: 201).

Byrne could likely offer a simpler solution to the difficulties listed above by endorsing the Fregean interpretation of v-propositions suggested earlier, though at some cost to the theory's advertised parsimony. But keep in mind that even with this Fregean account, the ban on treating v-propositions as visual representations would remain in effect, on pain of abandoning transparency altogether.[5]

3. The First Objection

According to Byrne, if you see that p_V, you must also believe that p_V (2012: 205-6). This is required by the inferentialist picture. To conclude, on the basis of SEE, that I see that p_V, I must accept the truth of the premise, p_V, at least if my act of concluding is itself supposed to take the form of a belief. Believing the conclusion simply on the basis of, for example, *entertaining* the premise or desiring it would not be an inference, whatever it was. But if I accept the truth of p_V, I believe p_V.[6]

---

[5] Also on pain of abandoning Fregeanism about visual content; see, for example (Chalmers 2006: 172).

[6] Note that this does not in any way contradict or presume to answer Valaris's (2011) objection to Byrne: that inference rules should work just as well in hypothetical or suppositional reasoning as in categorical reasoning. Presumably, Valaris would agree that if I am engaged in hypothetical reasoning, and so do not believe the premise, I need not believe the conclusion. The problem for Byrne which

A consequence of this is that if I know myself to be experiencing an optical illusion, I must have inconsistent beliefs (2012: 206). I see the Müller-Lyer lines and, being familiar with them, know them to be of equal length. But the illusion persists, so I also believe them to be of unequal length, or I believe a more complex v-proposition that entails that they are of unequal length. But there is nothing irrational about experiencing a persistent optical illusion.

It's natural to think that the Fregean interpretation of v-propositions would help. After all, one of the great advantages of Fregeanism is that it explains simply how beliefs with incompatible truth-conditions (e.g., *Hesperus is visible in the evening* and *Phosphorus is not visible in the evening*) can be jointly held without irrationality. This does not help with the victim of a persistent optical illusion, however, at least not in those cases in which *she herself* regards the visual experience as illusory. Even if we grant a Fregean interpretation, the consciously illuded subject would be at best like the person who believes *Hesperus is visible in the evening,* believes *Phosphorus is not visible in the evening*, but also believes that *Hesperus is Phosphorus*; or she would be like someone who believes *Phosphorus is not visible in the evening* while simultaneously believing *that belief of mine about Phosphorus is false.* Byrne's commitment to the irrationality of consciously illuded subjects seems unavoidable.

In response to this objection, Byrne writes:

> It will not do simply to claim that the illuded subject is not, or need not be, irrational. Taken as a claim about a rational ideal, its truth is not evident. Taken as an ordinary sort of remark, on the other hand, it is true but not in conflict with belief-dependence [of perception]. The belief that the subject knows to be false (e.g., a certain v-proposition that

---

Valaris identifies is that if I assume for the sake of argument that it is raining, it does not seem that I need to assume for the sake of argument that I believe it is raining.

is true only if the lines are unequal) does not influence her verbal reports about the lengths of the lines, or any plans for action based on the lengths of the lines. She is not therefore 'irrational' in the practical sense of an ordinary accusation of irrationality. … (2012: fn. 38)

Byrne's position is thus that the subject of persistent optical illusion may be irrational in the strict sense, but we do not ordinarily think of her as so, because her belief in the relevant v-proposition is effectively partitioned off from the rest of her beliefs.

How does this partitioning work? On the one hand, the belief survives the judgment that it is false. Byrne compares this with the modularity that delusional beliefs may enjoy—often surviving a person's recognition of their delusional status (2012: 206). At the same time, Byrne explicitly states that the belief in question will not be expressed in assertions or bring about intentions. Presumably the agent will also refrain from drawing inferences on its basis: since the v-proposition the agent believes contradicts another proposition she believes, application of standard inference rules would lead to explosion in her beliefs. Furthermore, the requirement that she not form intentions on the basis of the belief will only be met if she does not draw inferences on its basis (or those inferences are themselves partitioned). The partitioning, then, comes on both sides: the belief does not respond to its standard inputs, such as the belief that it is false, nor does it have any of the standard outputs: assertions, intentions, or further beliefs.

One obvious and troubling consequence is that the belief does not seem to play the functional role of a belief. It is a very standard commitment, however, that a given attitude-type possesses its functional properties necessarily. (Note that this is weaker than functionalism about mental states; it does not claim that mental states are functional states, that they have their particular mental properties in virtue of their functional properties, or that functional properties are sufficient to determine mental properties—it

10

simply claims that some state is, for example, a belief only if it has the functional properties of a belief.)[7]

If Byrne wishes to avoid a highly revisionary ontology of mental states, then, he must think that the functional properties are masked rather than missing. Some evidence that this is his position is suggested by this argument in favor of identifying perceptual states with beliefs:

> …[P]resumably some animals with visual systems very similar to ours (some other primates, say), cannot cognitively override visual illusions: in this sense, for them, seeing is always believing. Belief is thus built into their visual systems. And since we have basically the same visual systems, seeing is believing for us too. (2012: 205)

Now the fact that certain primates cannot override the deliverances of their visual systems, whereas humans can, might suggest that those deliverances play different functional roles in the two kinds of psychologies. But Byrne's idea seems to be that perceptual states had the functional role of beliefs in distant evolutionary ancestors, and that this gives some reason to think that the appropriate functional properties are still latent in our perceptual states.

Byrne can insist then that visual states have the functional properties of beliefs, holding only that these properties fail, for some reason, to manifest. He can insist that the relevant functions are *masked* rather than absent altogether. I don't wish to claim that this is an unreasonable position. What is objectionable, rather, is that no independent

---

[7] The claim could even be further weakened: to be a belief a state must have *most* or *enough* of the functional properties of belief. The concern is that Byrne's perceptual beliefs have near to none.

justification for this interpretation rather than the rival has been provided—that is, he has not provided any independent reason why we should regard visual states in adult humans as possessing "belief-y" dispositions that are masked, as opposed to lacking the dispositions altogether. (His point that "we have basically the same visual systems" is a *non-sequitur* in this context; the question is whether those visual systems play the same functional role in our psychologies—and if *we* can cognitively override their deliverances, there is substantial reason to think not.) But this is to say that Byrne's account depends on a contentious metaphysical commitment, which one would have no reason otherwise to make, and that is exactly the sort of thing that makes theories uneconomical.

This leads to the basic problem with Byrne's account: the proposed partitioning of the belief is extraordinarily *ad hoc*, especially in its details. To begin with, Byrne has proposed the existence of a belief that does not go away even after its possessor judges it false, simply because that is what his theory demands. There is no independent reason to accept the presence of such beliefs. The belief must also be partitioned, so that its possessor does not assert, intend, or infer on its basis. But again, the only reason to accept such a partitioning is to answer evidence that appears to disconfirm the theory— namely, that people do not act or reason as though they had such beliefs. This is already objectionably *ad hoc*. There is also the question of why such extensive partitioning does not undermine the state's status as a belief. Byrne is correct that partitioning *may* merely mask the relevant functional properties rather than eliminating them. But to go further and say that this is so is to insist on a particular thesis about the metaphysics of dispositions and functional properties in the absence of any argument in its favor—other than that is what inferentialism demands. Even more striking, however, is that the partitioning is not absolute. There will be one exception to the general walling-off of the illusional-belief with respect to inferences: the illuded subject will still apply inference

rule SEE to it. In other words, the one exception to this general partitioning is, by remarkable coincidence, the one required for the theory of self-knowledge to work in the first place. Even more remarkably, this pattern will be repeated, with the exact appropriate exemption, for the beliefs of every sense modality.[8]

The main case Byrne offers in favor of his inferentialist account of self-knowledge is, as he puts, *economy*. But this economical epistemology is purchased through profligacy in the theory of mind: with *ad hoc* partitionings, further *ad hoc* exemptions to those same partitionings, and *ad hoc* commitments on the metaphysics of functional properties. So the theory has not really achieved any sort of theoretical economy. It has simply moved several important lines of its budget onto another theoretical ledger.

4. Intention

A transparency-based account of knowledge of one's intentions must identify some fact about how the world is, from the agent's point of view, which indicates to the agent what it is that he or she intends. In the case of Byrne's inferentialist account, this means the relevant inference rule must specify the type of proposition that can serve as premise.

---

[8] A referee wonders if Byrne's partitionings are really that *ad hoc*. After all, we know the visual system to be relatively encapsulated. Couldn't Byrne explain the specific partitioning of visual beliefs in terms of this general encapsulation of the visual system? Possibly, but this would still fail to address two of the key ways in which the account is *ad hoc*. First, why don't these partitionings undermine the state's status as a belief? If Byrne wishes to say that it still manifests *enough* of the functional properties to count as a belief, the burden is on him to state what these manifestations are, given his claim that the state "does not influence [the illuded subject's] verbal reports about the lengths of the lines, or any plans for action based on the lengths of the lines." On the other hand, if Byrne grants (as he seems to) that the *manifestations* are lacking, he must explain why we should regard the functional properties as present, but masked or otherwise latent, rather than absent altogether. To insist on masking without independent grounds is *ad hoc*. Second, he must offer some independent grounds why SEE would be an exception to the visual system's general encapsulation. At present, nothing has been offered, and the only justification seems to be that inferentialism requires that it is so.

One obvious answer is that I determine what I intend to do by looking out at the world and figuring out what I have most reason to do (cf. Moran 2001). This might lead us to an inference rule such as:

OUGHT: If you ought to ϕ, believe that you intend to ϕ!

But such a principle could not be self-verifying, since we can be akratic (2011c: 214). What's more, there will be situations in which the reasons are neutral between several options, but in situations such as these, we still know which of the several acceptable options is the one we actually intend (Way 2007). Consequently, Byrne rejects the idea that we know our intentions by knowing those normative facts that would justify the intended action. Instead, he argues that when a person intends to ϕ she also believes that she will ϕ. A person thus knows her intentions on the basis of her belief that she will perform the intended action (2011c: 215ff.).

We need, then, a way to distinguish mere predictions about what one will do in the future from intentions. (I may predict that I will eat too much cake at the party because I always do, without intending to eat so much cake.) So Byrne restricts the application of the inference rule to those cases in which the prediction is not based on evidence. He writes:

I know that I am going to fail the exam because I know that I am poorly prepared; I know that I will be wearing down my sneakers because I know that I will be wearing them when I run the marathon. That is, I know on the basis of evidence that I will fail the exam and wear down

my sneakers.

However, as Anscombe points out, sometimes one's knowledge of what one will do is not arrived at by these familiar means. … [T]hose present and future actions that can be known 'without observation' are those that one intends to perform: if I know without observation that I will fail the exam, I intend to fail the exam; if I know without observation that I will run in the marathon tomorrow, I intend to run in the marathon tomorrow. (2011c: 218)

Byrne adds that Anscombe's phrase "knowledge without observation" is not strictly speaking accurate, and offers the replacement "knowledge not resting on evidence" (ibid.). So we know our intentions on the basis of our beliefs about what we will do in the future, along with the fact that these beliefs are not based on evidence.

The actual inference rule offered by Byrne does not mention anything about evidence, however; this rule, which he calls the "bouletic schema," takes the following simple form:

I will φ
_____
I intend to φ

(2011c: 216)

The restriction to cases in which belief in the premise is not based on evidence is presented in terms of conditions in which the bouletic schema is *defeasible*. Byrne writes:

"One will not reason in accord with the bouletic schema if one believes that one's belief that one will φ rests on good evidence that one will φ" (2011c: 218).

Once again, Byrne argues this rule, *given awareness of the defeating conditions*, will be *practically* strongly self-verifying:

> Privileged access is explained because the bouletic schema is *practically* strongly self-verifying: for the most part, if one reasons in accord with the schema (and is mindful of defeating conditions, for instance the one just noted), then one will arrive at a true belief about one's intention. (2011c: 219)

So we are now owed an account of how we are able to mind those defeating conditions.

5. The Second Objection

I can have evidence that *p* and believe that *p,* and yet fail to believe that *p* on the basis of evidence. I may have good evidence that lowering taxes would reduce unemployment; but the actual basis of my belief may simply be my desire to pay less in taxes. Consequently, there are two ways of specifying Byrne's defeating condition. Byrne could hold that a person should refrain from reasoning according to the bouletic schema when she believes she has sufficient evidence for her belief that she will *φ*; or he could hold that a person should refrain from reasoning in according with the bouletic schema when she believes that her belief that she will *φ* is *based on* sufficient evidence that she will *φ*.

Consider the first interpretation: the mere presence of sufficient evidence that I will *φ* makes the judgment that I intend to *φ* unwarranted. But for most of the actions I

intend, I will have sufficient evidence that I will perform the action. Most of my actions are not akratic, however it may sometimes feel. So most of the time, if I intend to φ, I believe the reasons favor φ-ing. Since I know myself to be rational *for the most part*, if I believe that the reasons favor φ-ing, I must think there is good evidence that I will φ. What, after all, would a generally rational person do?

Even if we agree not to count evidence on the basis of perceived reasons (for the sake of argument—the author has a hard time seeing any justification), inductive evidence often surrounds our intentions—we are predictable creatures. Two akratics could have the same evidence (the same track record of failure), and each could believe of himself that he will eat too much cake at the party (cf. Ashwell 2013: 253ff). But one could also know himself not to currently intend to eat so much, though he knows such an intention will form at some point after his arrival; whereas the other could know himself to be already akratically intending the impending gluttony. But if their evidence is the same, there must be some explanation why it defeats the inference in the first case but not the second.

This takes us to the second interpretation, which would allow us to accommodate the two akratics and the other cases as well. But now if we are to be "mindful of the defeating conditions" we must know not only what we believe, but why we believe it. We must know that we have the belief *because* of that evidential support. But that requires we know, not only what our beliefs are, but, for example, what causal relations hold between our beliefs. Or, if one is unhappy with causal accounts of believing for a reason, we must know some other relation to hold between the beliefs: the counterfactuals under which we would and would not continue to hold the beliefs, or

which primitive relations of *believing on the basis of* hold among our beliefs. Whatever the precise relation, we are entitled to ask how we know that.

Even independently of Byrne's story about intention, it seems to be part of common sense that people generally know not just what they believe, but why they believe it. Of course, in individual cases we might be self-deceived. But any theory of self-knowledge not aiming at significant revision should predict that we *generally* know not only our attitudes, but our reasons for holding the attitudes.

We could know inferentially, but Byrne has not provided any inference rule that provides knowledge about the causal, counterfactual, or primitively evidential relations that hold between one's beliefs. BEL only leads to conclusions about what one's beliefs are, not how they interact with one another. There is, moreover, a general reason to think that there could not be such an inference rule, given Byrne's existing commitments. Such an inference rule would need to tell us which relations one's beliefs stand in to one another. But consider what the example of the two akratics shows us. Both akratics have beliefs with the same content. But despite identical content, those beliefs stand in different relations to one another: the first akratic believes he will eat too much cake on the basis of believed inductive evidence; the second believes the inductive evidence, and believes he intends to eat too much, but the second belief is held "not on the basis of evidence," at least if it is the type of Anscombean knowledge Byrne's account depends on.

Byrne's inferentialist account of knowledge of intention unfortunately *precludes* any inferentialist account of knowledge of the relations of support among beliefs (and since on his account, knowledge of former depends on knowledge of latter, it cannot work). The simple fact an agent has some set of beliefs underdetermines which relations hold between them, even if we restrict ourselves to very ordinary cases. Sue, for

example, believes (1) *All crows are black*, (2) *That bird is a crow* and (3) *That bird is black.* But this tells us nothing about the evidential relations holding between those beliefs. She might have reasoned deductively—from (1) and (2) to (3). She might have reasoned inductively to (1), with (2) and (3) jointly forming an instance supporting the generalization. She might have reasoned abductively, concluding that that bird is a crow rather than *sui generis* black scavenger, in which case (1) and (3) form part of (2)'s support. Those relations that hold between contents—for example, that p and q jointly entail r— also underdetermine the relations that hold between the beliefs with those contents, for the simple reason that neither inductive nor abductive justification tracks entailment at all. But this means that one cannot reason from the relations between the contents to the relations of support among the beliefs: inductive and abductive reasoning don't track the relations among the contents. Nor can one reason from the presence of the attitudes themselves to their relations of support.

In order to have inferentialist knowledge of the basis of one's belief, then, one would presumably need to infer from a combination of one's beliefs, and one's judgments about the evidential situation (where that presents itself at least as an objective feature of the world—in order to respect the transparency intuition).[9] I will not canvass possible inference rules taking both one's beliefs and one's assessment of one's evidential situation as inputs—only note that they may get quite complicated in order to achieve self-verification. For my purposes it is enough to show that if Byrne is right, and there is the possibility of Anscombean knowledge "not on the basis of evidence," no such rule could be self-verifying, at least with respect to my beliefs about my own future actions. After all, that I have decisive evidence that I will φ doesn't tell me if I believe that I will φ *because* of that evidence or because of my intention (the two akratics showed us this).

---

[9] Thanks to an anonymous referee for suggesting this.

So, given Byrne's commitments, knowledge of whether my belief is held on evidential grounds cannot be on the basis of an inference rule. So what is the basis? Perhaps we know on the basis of third-personal evidence. Remember, though, the question is whether I believe *on the basis of evidence* that I will *φ*. But what about my behavior, or past history, or generalizations about human nature would answer that question for me? It's hard to think of anything. It's very hard to think of anything that will be present in all of the cases in which, according to common sense, I know that I have an intention.

But if it isn't inference rules, and it isn't third-personal forms of evidence, what could it be, except some uniquely introspective epistemic capacity? This would mean, though, that Byrne's account completely fails to explain "self-knowledge in terms of epistemic capacities and abilities that are needed for other subject matters."

Could Byrne avoid difficulty here with a different inference rule—one that doesn't require the possibility of Anscombean knowledge if we're to mind defeaters? It's possible, but finding one will be tricky. (Ashwell 2013) suggests in passing that we could know our intentions on the basis of our desires, and goes on to offer an inferentialist account of knowledge of desires.[10] But just as I sometimes must form an intention in cases where the reasons favor both options equally, I must sometimes form an intention when I am indifferent between my options. It seems I can know what I intend in these cases. An account of self-knowledge must be able to handle knowledge of those intentions that are to some degree arbitrary.

6. The Moral?

---

[10] Ashwell also argues that this account is probably not economical in Byrne's sense.

So transparency can't be had for cheap. Admittedly, plenty of its advocates have found it compelling enough that they are willing to pay a high price. Richard Moran argues that the transparency of belief and other attitudes makes self-knowledge different in kind from other, philosophically more familiar forms of knowledge (2001 and 2012). In a recent paper, Markos Valaris argues that transparent self-knowledge is explained by a *sui generis* rational activity (2014: 10). Lauren Ashwell (2013) offers an inferentialist account of the transparency of desire which she frankly admits to be incompatible with considerations of economy. Of course, economy is a theoretical virtue—and if transparency accounts must give it up, that is a cost, even if it is compensated for elsewhere.

In any case, there is a wider moral: Byrne's difficulties illustrate two general challenges transparency-based accounts must address.

First, attitudes are not simply distinguished on the basis of their content, but also on the basis of how they relate (or relate the agent) to that content. To illustrate, a belief can have the same content as an intention (and so the bouletic schema risked confusing the two). What distinguishes the attitudes when their content is identical is how that content is presented: the belief as a way things are, the intention as an end or aim to be realized. The challenge for the transparency theorist is to explain how we come to know not just the content of our thoughts, but how our thoughts are related to that content, despite her commitment to the transparency of one of the *relata*.

The obvious answer is that you come to know (either inferentially or in some other way) your mental states on the basis of the other *relata,* the content (cf. Ashwell 2013; and Valaris 2014).[11] That may not be possible, however, with all of the relevant

---

[11] Both Ashwell and Valaris argue, for example, that desires in some way present their objects as valuable, allowing us to know our desires by asking what is desireable. This seems to commit them to

states, as we saw in the case of intention—especially intentions in cases when we are indifferent and judge our options equally reasonable; these must be made arbitrarily, which seems to mean, when there is no distinguishing feature in their content.

On the other hand, Byrne did distinguish perceptual states on the basis of content, by proposing categories of v-propositions and o-propositions and so on. Yet even with the special content, the account required multiple *ad hoc* stipulations to avoid incredible predictions. This points to a second, seemingly more difficult challenge. Perception relates to its content in a manner that does not *commit* the agent as belief does (cf. Schafer 2013). I can see the lines as unequal, without accepting that they are unequal; but if I believe them to be unequal, that just is accepting that they are unequal.

Byrne simply denied that perception was non-committal: seeing was literally believing. This is counterintuitive, but given his commitment to transparency, well-motivated. For as we said, in a transparency account, the obvious way we come to know the specific ways our mind relates to its content is on the basis of the content itself. But with non-committal states, the question is how I have access to the content, when I do not hold that content to be part of the world.[12] Why aren't both *relata* transparent (or simply invisible)?

A transparency-based inferentialist account of self-knowledge actually comes at the price of economy. It does so because of the difficulties of specifying, on the basis of those propositions to which the agent is committed, the variety of attitudes she may hold to that same content. Non-committing states create further problems for it: to explain

---

a *guise of the good* position on motivation, and in the case of Ashwell, a version especially similar to that found in (Johnston 2001; and Oddie 2005; but also see Tenenbaum 2007; and Schafer 2013). For examples of recent criticisms of such views, see (Schroeder 2008; and Baker 2014).

[12] (Valaris 2014: 15) suggests that this could perhaps be explained if perceptual states take non-propositional content. The suggestion is intriguing, but it would need to be developed in more detail before its promise could be assessed; and we would need to know if this solution could be extended to other plausibly non-committal states, such as, say, imagining.

access to the content that provides knowledge of these states, it must make counterintuitive claims about commitment. Some other variety of transparency may be better able to solve these problems, but they are problems for which a solution must be given.

References

Anscombe, G.E.M. 1957. *Intention.* Blackwell University Press.

Ashwell, L. 2013. Deep, Dark… or Transparent: Knowing Our Desires. *Philosophical Studies* 165: 245-56.

Baker, D. 2014. The Abductive Case for Humeanism over Quasi-Perceptual Theories of Desire. *The Journal of Ethics and Social Philosophy* 8: 1-29.

Boyle, M.   2011. Transparent Self-knowledge. *Proceedings of the Aristotelian Society, Supplementary Volume* 85: 223-41.

Byrne, A. 2005. Introspection. *Philosophical Topics* 33: 79-104.

_____. 2011a. Knowing That I Am Thinking. In *Self-Knowledge,* ed. A. Hatzimoysis, 105-24. Oxford: Oxford University Press.

_____. 2011b. Knowing What I Want. In *Consciousness of the Self: New Essays,* eds. J. Liu and J. Perry. Cambridge: Cambridge University Press.

_____. 2011c. Transparency, Belief, and Intention. *Proceedings of the Aristotelian Society, Supplementary Volume* 85: 201-19.

_____. 2012. Knowing What I See. In (Stoljar and Smithies 2012: 183-210).

Carruthers, P. 2013. *The Opacity of Mind: An Integrative Theory of Self-Knowledge,* Oxford: Oxford University Press.

Chalmers, D. 2006. The Representational Character of Experience. In *The Future for Philosophy,* ed. B. Leiter, 153-81. Oxford: Oxford University Press.

Evans, G. 1982. *The Varieties of Reference.* Oxford: Oxford University Press.

Humberstone, L. 1992. Direction of Fit. *Mind* 101: 59-83.

Johnston, M. 2001. The Authority of Affect. *Philosophy and Phenomenological Research* 63: 181-214.

Lycan, W.G. 1996. *Consciousness and Experience,* Cambridge, MA: MIT Press.

McGeer, V. 1996. Is 'Self-knowledge' an Empirical Problem? Renegotiating the Space

    of Philosophical Explanation. *Journal of Philosophy* 93: 483-515.

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

_____. 2012. Self-knowledge, 'Transparency', and the Forms of Activity. In (Stoljar

    and Smithies 2012: 211-38).

Oddie, G. 2005. *Value, Reality, and Desire.* Oxford: Oxford University Press.

Schroeder. M. 2008. How Does the Good Appear to Us? *Social Theory and Practice* 34:

    119-30.

Schafer, K. 2013. Perception and the Rational Force of Desire. *The Journal of Philosophy*

    110: 258-81.

Shoemaker, S. 2012. Self-intimation and Higher-order Belief. In (Stoljar and Smithies

    2012: 239-58).

Stoljar, D., and D. Smithies. 2012. *Introspection and Consciousness.* Oxford: Oxford

    University Press.

Tenenbaum, S. 2007. *Appearances of the Good: An Essay on Practical Reason.* Oxford:

    Oxford University Press.

Valaris, M. 2011. Transparency as Inference: a Reply to Byrne. *Proceedings of the*

    Aristotelian Society, Supplementary Volume 111: 319-24.

_____. 2014. Self-knowledge and the Phenomenology of Belief. *Philosophers' Imprint*

    14(8): 1-17.

Way, J. 2007. Self-knowledge and the Limits of Transparency. *Analysis* 67: 223-30.