II—Matthew Boyle

Transparent Self-Knowledge

I distinguish two ways of explaining our capacity for 'transparent' knowledge of our own present beliefs, perceptions, and intentions: an *inferential* and a *reflective* approach. Alex Byrne (2011) has defended an inferential approach, but I argue that this approach faces a basic difficulty, and that a reflective approach avoids the difficulty. I conclude with a brief sketch and defence of a reflective approach to our transparent self-knowledge, and I show how this approach is connected with the thesis that we must distinguish between a kind of self-knowledge that is of oneself as agent and another kind that is of oneself as patient.

> Let not the mind, therefore, seek itself as if it were absent, but let it take care to discern itself as present.
>
> —Augustine, *De Trinitate*, 10.9.12

I

Over the past few decades, and especially in the last ten years or so, the idea that our standpoint on our own mental lives is in some sense 'transparent' to our standpoint on the world at large has played an increasingly prominent role in philosophical discussions of self-knowledge. This idea has inspired important work on how we know ourselves to hold 'judgement-sensitive' attitudes such as belief, desire and intention,[1] and it has also provided the impetus for a reconsideration of our knowledge of what our sensory and perceptual experiences are like.[2] In Alex Byrne's recent work (2005, 2011), these two tendencies are merged into a compact and unified account of how we know our own minds. The account is in some respects deflationary: if Byrne is right, the fact that we have privileged access to the contents of our own minds has less to teach us about the nature of our men-

[1] See, for instance, Evans (1982), Dretske (1994), Gallois (1996), Moran (2001) and Byrne (2005).

[2] See, for instance, Harman (1990), Shoemaker (1994) and Tye (2000).

Proceedings of the Aristotelian Society Supplementary Volume lxxxv
doi: 10.1111/j.1467-8349.2011.00204.x

tality than some authors have supposed. But Byrne's project is also ambitious: he aims to show that essentially the same cognitive capacities underlie the whole range of things we can know about our own mental states, and that these capacities involve nothing exotic—no special power of inner sense or self-determination—but merely a special deployment of powers possessed by anyone who can draw inferences about any topic whatsoever.

My own contribution to this subject consists of an essay (Boyle 2009) defending an idea of Richard Moran's about how we know our own judgement-sensitive attitudes. Moran (2001) proposed that the reason why I can normally answer the mind-focused question whether I believe that *P* 'transparently', by reflecting on the world-oriented question whether *P*, is that my reflection on the latter topic is an exercise of my capacity to 'make up my mind', so that my knowledge of my belief concerning *P* is normally grounded in my capacity actively to determine whether to believe the proposition in question. He went on to suggest that this sort of account might, with suitable modifications, be extended to other species of judgement-sensitive attitudes.

My essay defended Moran's position against a criticism brought by several authors: that his account cannot be correct, since there are many kinds of mental states (e.g. feeling toothache, seeing what appears to be a dagger before me) of which we have privileged self-knowledge, but to which an account of Moran's sort plainly does not apply. I granted—as Moran himself does—that there are states of this kind, but I argued that there is no good reason to assume that an account of self-knowledge must explain all of our self-knowledge in essentially the same way (I called this the 'Uniformity Assumption'), and I tried to show that there is in fact reason to think that the kind of self-knowledge Moran identifies must have a fundamental place in an adequate theory of self-knowledge. I suggested that Moran's insight sheds light on an idea famously associated with Kant: that we must distinguish between an active aspect of self-knowledge that is knowledge of ourselves as spontaneous beings and a passive aspect grounded in our power of sensible receptivity.

Byrne (2011) attacks these claims from a direction I did not anticipate. The objectors I was trying to answer generally did not make much of the phenomenon of transparency: they held that the fact that our answers to questions about our own mental states are systematically correlated with our answers to questions about the

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

world at large is, if anything, a by-product of our normal way of knowing our own minds, not the key to an account of this knowledge. According to Byrne, however, precisely this correlation is the basis of our self-knowledge, and he thinks this holds true across the board. He thus argues that there *is* good reason to suppose that a uniform account of self-knowledge is possible. Moreover, he denies that the capacity to make up one's mind plays an essential role even in the account of our transparent knowledge of our judgement-sensitive attitudes. My paper appears in his discussion mainly as a foil to his own view on these two points.

I am grateful to Byrne for the attention he has given to my paper, and for the opportunity he has given me to revisit these issues. I think his emphasis on the generality of the phenomenon of transparency brings out something important, something I did not sufficiently appreciate when I first wrote on this topic. But I also think that Byrne's strategy for explaining transparency is wrong-headed in a fundamental way, and that a sound explanation of the matter will not eradicate the distinction between a kind of self-knowledge that is of myself as agent and another kind that is of myself as patient. Rather than repeating my earlier arguments, I will try to approach these issues from a new angle. I will begin by contrasting an *inferential* and a *reflective* approach to explaining transparency (§II). Byrne advocates an inferential approach, but I will argue that this approach cannot make our normal capacity for self-knowledge intelligible (§III), and I will then explain why I think a reflective approach can do better, and what place a distinction between agency and patiency has in such an account (§IV).

II

What is 'transparency', and what sort of explanation do we need of it? As Byrne notes, the term has been used to mark a number of different points: the fact, famously noted by Evans (1982), that I can normally answer the question whether I believe *P* simply by considering whether *P*; but also the fact, famously noted by Moore (1903), that if I attempt to focus my attention on the intrinsic quality of 'my sensation of blue', all I can discover is blueness, a property of worldly things. Byrne thinks these points have something in common, and I agree. We can express the commonality, vaguely but

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

evocatively, as follows: I can know various aspects of the nature, content and character of my own mental states by attending in the right way, not to anything 'inner' or psychological, but to aspects of the world at large.[3] Indeed, it seems that, for various sorts of mental states, there is in the normal case no other way to attend to them: all there is for me to contemplate in my sensation of blue is the (apparent) blueness of some worldly thing, and all there is for me to attend to in my belief that *P* is the (apparent) fact that *P*. Various questions about my own present mental state are thus normally 'transparent' for me to questions about the world at large.

Byrne focuses, in the first instance, on the transparency of questions about my own beliefs, and I will do likewise. The puzzling thing about this transparency is how the world-oriented reflection can bear on the question about my own mental state. Whether *P* and whether I believe *P* are, prima facie, quite different topics. It certainly does not seem to be a general truth that if *P*, then I believe *P*: I am not omniscient, after all. Yet my normal knowledge of what I believe does seem to rest in some way on my capacity to consider how things stand in the world at large.

At this point, we should distinguish two reactions to these observations. A *credulous* reaction accepts that we do somehow draw a conclusion about our state of mind from a fact about the world. An *incredulous* reaction holds that this cannot be right, and that the appearance that we make such a transition must reflect some misunderstanding. Byrne reacts credulously; indeed, his formulation of the puzzle of transparency presupposes that this is how we should react (cf. Byrne 2011, pp. 203–4). For my part, I am incredulous. I will give some reasons for my incredulity in the next section, but first I want to consider how a philosopher who reacts in each of these two ways might explain the transparency of belief.

For the philosopher who reacts credulously, the puzzle of transparency will consist in the fact that our normal knowledge of what we believe rests on a cognitive transition whose reasonableness is hard to understand, and the task will be to explain how the transition can be reasonable. This is Byrne's approach: he grants that a subject who transparently forms a view about what she believes

[3] Part of what is left vague in this formula is the relation between the world-oriented and the mind-focused thinking: as Byrne notes (2011, p. 203), the nature of the step from a thought about the world to a thought about myself needs clarification. But it will turn out that how to clarify this is the crux of our dispute, so I will leave the matter vague for now.

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

makes a transition between contents that are not deductively or evidentially related to one another, but he argues that, nevertheless, it is in context a perfectly good transition to make, since a subject who infers *I believe P* from *P* must, inasmuch as she accepts the premiss, in fact believe *P*. In earlier work, Byrne (2005) characterized such a subject as 'following a good epistemic rule'; here he simply says that she makes a good inference. Some might hesitate over this label, but if by 'inference' we simply mean a non-accidental transition between belief contents, where the reasonableness of the transition is open to assessment, it should be unobjectionable to call this as an *inferential* approach of transparency.

A philosopher who reacts incredulously, by contrast, will hold that only a madman could draw such an inference, that it won't do to hold that normal self-aware believers are mad, and hence that the inferential approach to transparency cannot be correct. He will therefore need to offer some other interpretation of doxastic transparency, one that does not represent us as drawing conclusions about our own minds from facts about the world. Yet if he is not to deny the phenomenon of transparency altogether, he must recognize *something* right in the idea that our knowledge of what we believe is grounded in our capacity to consider how things stand in the world at large. How can he do this?

He can do so as follows. Instead of thinking of the subject as making an inference from *P* to *I believe P*, he can think of the subject as taking a different sort of step, from *believing P* to *reflectively judging* (i.e. consciously thinking to himself): I believe *P*. The step, in other words, will not be an inferential transition between *contents*, but a coming to explicit acknowledgment of a *condition* of which one is already tacitly aware. The traditional philosophical term for this sort of cognitive step is 'reflection', so I will call this a *reflective* approach to explaining transparency.[4]

A comparison will bring this proposal into sharper focus. Consider a person who *asserts P* by saying, in an assertive tone of voice,

[4] Thus Locke speaks of reflection as 'the notice which the Mind takes of its own Operations, and the manner of them', a notice that gives us ideas such as 'Perception, Thinking, Doubting, Believing, Reasoning, Knowing, Willing, and all the different actings of our own Minds' (Locke 1975, II.i.4). Compare also Kant's definition of reflection as 'the consciousness of the relation of given representations to our various sources of cognition' (Kant 1997, A260/B316). Locke and Kant have quite different conceptions of the nature of the reflective act, of course, but they agree in supposing that it involves turning our attention from what we are representing to the mode of our own activity in representing it.

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

'*P*!' Such a person might give no thought to the nature of his own speech-act: he might be engaged in heated debate, with his attention focused wholly on the matter in dispute. Nevertheless, there is a sense, palpable if difficult to explicate, in which he will normally know what he is up to—that he is not inquiring whether *P*, or positing *P* for the sake of argument, but *claiming P*. Furthermore, if he has mastered the needed vocabulary, he will on reflection be able to repeat his act in a way that makes its nature explicit—for instance, by saying 'I am claiming: *P*!' How did he know that this is what he was doing? It would be absurd, I think, to suggest that he inferred it from the apparent truth of the *content*: *P*. The right account, surely, is that he all along had some sort of tacit awareness of what he was up to, and that this awareness, together with his mastery of certain classificatory expressions, allowed him to make his activity explicit, for us and for himself. All he had to do was to turn his attention to what he was already knowingly doing—i.e. to reflect.

The reflective approach explains doxastic transparency in a similar way: as a matter, not of inferring of a psychological fact from a fact about the world, but of shifting one's attention from the world with which one is engaged to one's engagement with it—an engagement of which one was already tacitly cognizant even when one's attention was 'directed outward'. The reflective approach thus does not seek to explain *how we acquire* doxastic self-knowledge. It explains this knowledge, not by appeal to some mechanism or method that allows the subject to know an otherwise unknown fact about himself, but in terms of the nature of belief itself. It treats the following as a basic, irreducible fact about believing as it occurs in a creature capable of reflection: a subject in this condition is such as to be tacitly cognizant of being in this condition. Hence, in the normal and basic case, believing *P* and knowing oneself to believe *P* are not two cognitive states; they are two aspects of *one* cognitive state—the state, as we might put it, of knowingly believing *P*.[5]

[5] Note that these remarks are all subject to the qualification 'in a creature capable of reflection'. I do not deny that creatures incapable of reflection can have beliefs, and I take no position on whether their believing involves tacit knowledge of believing. I will shortly say more about the notion of tacit knowledge, and about the significance of the other qualification here: 'in the normal and basic case'.

For other versions of the idea that believing and knowing oneself to believe are normally not two cognitive conditions but one, see Moran (2001, pp. 30–1), Rödl (2007) and Shoemaker (2009). (Such views are often called 'constitutivist'. I avoid this label because it is sometimes taken to imply that one's judging oneself to believe *P makes it the case* that one

doi: 10.1111/j.1467-8349.2011.00204.x

This still leaves room for the reflectivist to recognize an important truth underlying the phenomenon of transparency. On his view, the important truth is this: the very same actualization of my cognitive powers that is my believing *P* is, under another aspect, my tacitly knowing that I believe *P*. Hence, to pass from believing *P* to judging I believe *P*, all I need to do is reflect—i.e. attend to and articulate what I already know. Something broadly similar will hold for other psychological conditions of which I can have transparent self-knowledge.

I will defer the task of arguing for this approach until later, but let me try to head off some objections it may provoke. Note, first of all, that this view need not imply that a subject is infallible in his convictions about his own beliefs (that whenever a subject believes himself to believe *P*, he must actually believe *P*). The proposal is that believing *P* involves tacitly knowing oneself to believe *P*, but this is consistent with one's sometimes believing oneself to believe *P* when one does not.

Many will worry, however, that this proposal attributes to subjects an implausible omniscience about their own beliefs (implying that whenever one does believe *P*, one knows oneself to believe *P*).[6] Are there not occasions on which we hold a belief (e.g. an unacceptable prejudice, an anxiety-inducing doubt, a realization we 'cannot face') but are not conscious of its existence, or are conscious of it only in an 'alienated' way, as a conviction which evidently governs aspects of our behaviour, but which we do not consciously endorse? I agree that there are such beliefs, but someone who holds that believing *P* and tacitly knowing oneself to believe *P* are two aspects of a single condition need *not* deny their existence. His view will simply be that when a belief is present but not consciously accessible, *so too is the knowledge of that belief*. This exploits a possibility that characterizes all knowing. Knowing that *P* is a standing state, one that puts its subject in a position, other things equal, to think knowledgeably that *P*. But it is a familiar fact that the availability of knowledge to conscious thought can be interfered with in various ways: distraction, confusion, or temporary inhibition of memory can put us out of touch with what we know; and no doubt things

believes *P*, but that is not my view; and anyway the term 'reflectivist' better captures the specific character of the account I defend here.)

[6] Compare, for example, Martin (1998, §II).

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

we know can be blocked from consciousness in more radical ways by repression, self-deception, and other distorting psychological tendencies. Thus it may be true that a person's believing *P* involves his knowing himself to believe *P*, and yet that a person can believe *P* without being conscious of it. This is *not* a paradox, so long as we are careful about the difference between being known and being accessible to conscious reflection. What is known is accessible to conscious reflection, *other things being equal*, but other things are not always equal.

So far I have just been sketching two approaches to transparency: inferential and reflective. In the next section, I will give some reasons for my incredulous reaction to Byrne's inferentialism. In the subsequent section, I will make a case for the reflective approach.

III

The basic reason to reject the idea that I infer a fact about my own psychology from a fact about the world is just this: the inference is mad. Byrne himself acknowledges a prima facie difficulty here (2011, p. 204), but I believe the problem is more serious than he recognizes.

Suppose for the sake of argument that I have arrived at the belief that I believe *P* by inferring according to Byrne's 'doxastic schema':

$$\frac{P}{\text{I believe } P}$$

To believe that I believe *P* is to hold it *true* that I believe *P*. Being a reflective person, I can ask myself what grounds I have for holding this true. The answer '*P*' is obviously irrelevant. I am asking what shows that the proposition *I believe P* is true, and a modicum of rational insight will inform me that, even if it is true that *P*, this by itself has no tendency to show that I believe it. What would support my conclusion, of course, is the fact that I, the maker of this inference, accept the premiss that *P*. But to represent *that* as my basis would be to presuppose that I already know my own mind on the matter, and that would undermine Byrne's account.

Byrne assumes that we at least *can* draw inferences according to the doxastic schema, but I think even this is not obvious. His idea, I

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

take it, is that an inferential disposition is simply a reliable tendency to pass from one believed content to another, and that for any ordered pair of contents, it is possible in principle to imagine a subject who has an inferential disposition to pass from the former to the latter, though some of these subjects will be irrational. I think this presupposes an unacceptable conception of belief and its relation to inference. The matter needs a fuller discussion than I can give it here, but I can put my objection in a provisional way by saying: a belief, once formed, doesn't just sit there like a stone.[7] What I believe is what I hold true, and to hold something true is to be in a sustained condition of finding persuasive a certain view about what is the case. Even if we grant that a disposition to pass from one content to another could deposit various arbitrary beliefs in my mind, those beliefs would be unsustainable if I, understanding their contents, could see no reasonable basis for holding them true.

We should, moreover, question the idea that inference is merely a reliable process that deposits beliefs in my mind. A (personal-level) inference is not a mere transition from a stimulus to a response; it is a transition of whose terms I am cognizant, and whose occurrence depends on my—in the normal case: persistently—taking there to be an intelligible relation between these terms. This is what makes it possible for an inference to leave me with a sustainable belief: I can reflect on why I draw a certain conclusion, and when I do, I can see (what looks to me to be) a reason for it. It is hard to see how the premiss of Byrne's doxastic schema could supply me with a reason to draw its conclusion.[8]

Byrne's inferential approach to doxastic transparency thus appears to face a dilemma: it must either represent the subject as drawing a mad inference, or else must admit that her real basis for judging herself to believe *P* is not the sheer fact that *P*, but her tacit knowledge that she *believes P*. The second horn of this dilemma should be unacceptable to Byrne: embracing it would mean giving up on his project. But—I want to suggest—there is nothing intrinsically unacceptable about this option. To embrace it is simply to

[7] For further discussion, see Boyle (forthcoming).

[8] Unless the subject thought himself omniscient. But *that* possibility only reinforces my point: the assumption of omniscience *would* make this a prima facie reasonable transition, though the assumption itself is deeply unreasonable. My objection is that, absent some such mad assumption, Byrne's doxastic inference is not an inference a person could intelligibly draw.

Proceedings of the Aristotelian Society Supplementary Volume lxxxv
doi: 10.1111/j.1467-8349.2011.00204.x

adopt a reflective approach. A main task of reflectivism will be to clarify what our tacit knowledge of such states can amount to, and how it belongs to the nature of the relevant mental states. It is here, we shall see, that Moran's appeal to rational agency becomes relevant: it supplies an account of how *certain* kinds of mental states are, by their very nature, tacitly known. But, we shall also see, this account does not apply across the board: our tacit knowledge of other kinds of mental states needs a different account.

So far, I have only criticized Byrne's account of doxastic transparency, but I think the idea that self-knowledge is achieved by inferring psychological facts from non-psychological facts is no more plausible when we turn to perception, pain or intention. Indeed, I think it is even clearer in these cases that the true basis on which we make judgements about our mental states is one that presupposes tacit knowledge of those states.

Consider our knowledge of our own perceptions. Byrne (2011, pp. 203, 213) observes, following Evans, that I can typically make accurate judgements about how things perceptually appear to me by focusing my attention, not on my own state of mind, but on the world around me. Thus I can, for example, normally say how things look to me simply by prefixing 'It looks to me as if' to my answer to the question 'How do things stand in the space directly in front of me?'[9] Does this not show that I can reasonably draw conclusions about my mind from facts about the world? I do not deny that it shows an important connection between these two topics, but it is an oversimplification to think the connection is one-way. The information I draw on in answering the question 'How do things stand in the space in front of me?' is not merely abstract information I happen to have about the world; it is information I am cognizant of *under a specific mode of presentation*, a mode that comes out in the forms of words we naturally use to express it: '*This* cat is grey', '*There's* a tomato'. These forms of words do not, of course, make explicit reference to any psychological state, but it is not an accident that philosophers commonly call the relevant 'this' a 'perceptual demonstrative'. Perceptual information is presented in such a way that a subject who can competently express the content of her perceptual experience is already implicitly marking a distinction between infor-

[9] This is of course an oversimplification. I will, among other things, need to restrict my answer to the sorts of facts that vision supplies, and to set aside any extraneous information I may have about, for example, trickily placed mirrors, the shapes of sticks in water, etc.

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

mation that is *perceptually given to her here and now* and other ways of having information. That is the force of the perceptual 'this': it marks a mode of presentation of things and their properties that is made available precisely by perception.

So although a subject who makes the perceptually-based judgement *This cat is grey* does not make a judgement whose *content* concerns her own mental state, the *form* of her judgement implicitly bears on the nature of her mental state. In so judging, she presupposes that she is perceiving the cat in question, even if she never makes this presupposition explicit to herself—even, indeed, if she does not possess the psychological concepts that would allow her to make it explicit. By the same token, if she does possess the relevant concepts, she merely knows how to make explicit aspects of what she is already presupposing. She does not arrive at knowledge of one realm of facts by inference from another, epistemically independent realm of facts.[10]

Similar points apply to our transparent knowledge of our own sensations and intentions. I will only discuss the latter case. Byrne (2011, pp. 216, 218) proposes that I can infer *I intend to* ϕ from *I will* ϕ, provided that I do not take myself to believe that I will ϕ on the basis of good evidence. The oddity of this proposal will come out if we distinguish two uses of 'will ϕ': an *intention-based* use that implies that my ϕ-ing will be the realization of my intention to ϕ (we can reserve the present progressive 'I am going to ϕ' for this special sort of representation of the future) and a 'will' of *blank futurity* that applies whether my ϕ-ing realizes an intention or not. Now, I certainly grant that if I think that I will ϕ, where this is the intention-based 'will', then I can rightly judge *I intend to* ϕ. But once again, this transition does not explain my knowledge of what I intend, but rather presupposes it. Suppose, however, that I believe I will ϕ, where this is the 'will' of blank futurity, a 'will' that might link me to such ϕ's as 'trip', 'meet a tall dark stranger', 'throw up'. Might I not very easily believe such things, without good evidence, and yet have no intention whatsoever of bringing about these eventualities? To infer my own intention simply on this sort of basis

[10] In an earlier paper, Byrne characterizes rules for drawing conclusions about my own mental states as 'neutral' if their antecedents 'are not specified in terms of the rule follower's mental states' (Byrne 2005, p. 94). If I am right, this label is misleading: a ground can be neutral in this sense without being genuinely non-committal as to the nature of the subject's mental states.

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

would be reckless, except where my conviction in the blank future proposition is itself grounded in my conviction that I will ϕ in the intention-based sense.

Part of what is odd about the idea that I might infer propositions about my present intentions from blank future propositions about myself is that it seems to get matters backwards. In certain instances, it seems, I believe that I will ϕ precisely *because* I (knowingly) intend to ϕ. My intention is a kind of commitment to ϕ, and my knowing acceptance of this commitment, together with my understanding of my own active powers, is the *ground* of my conviction that I will ϕ. The idea that the line of epistemic dependence runs in the other direction expresses a profoundly *alienated* picture of my knowledge of my own intentions—as if I must conclude to my own commitment to ϕ from an unaccountable inkling about what I will in fact do.[11] A subject who had to discover her intentions in this way would at best know *of* her own intentions; she would not know them through *seeing herself in* them. For her knowledge that she intended to ϕ would not be grounded in her knowing commitment to ϕ-ing. She would not know her intentions through seeing certain things as to-be-done.[12]

IV

I have already made a preliminary case for a reflective approach to our knowledge of our own beliefs, perceptions, and intentions. I now want to consider in a more general way what is at stake in this approach, and how it is connected with the contrast between agency and patiency.

[11] I am not suggesting that one must think explicitly about one's own intentions to determine what one will do. One can normally just consider one's options and make a decision about what one will do. But this is not the 'will' of blank futurity, but a specifically intention-based representation of what I *am hereby going* to do.

[12] See Byrne (2005, pp. 86–7) for a reply to the charge that his account implies alienated self-knowledge. The general idea is: I stand in an alienated relation to a mental state *M* only if *M* fails to play its normal inferential/causal role, so the mere fact that I *know* that I am in *M*, by whatever method, cannot entail that I am alienated from *M*. I think this misconstrues the objection. The concern is not that a subject who knows his mind in Byrne's way must experience some disturbance in the operation of his first-order mental states, but that his supposed *knowledge* of his mental states would not stand in the right relation to his being in those states. He would merely know *of* his own first-order states; he would not see himself *in* them.

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

The reflectivist rejects an explanatory demand that many theorists of self-knowledge accept. He denies that, in the normal, non-alienated case, being in a given mental state *M* and believing oneself to be in *M* are two distinct psychological conditions, and consequently denies that the task of a theory of self-knowledge is to explain how these conditions come to stand in a relation that makes the latter knowledge of the former.[13] We can call any account that accepts this conception of the task of a theory of self-knowledge an *epistemic approach*. This approach is shared by many otherwise diverse theories: those that posit a faculty of inner sense, those that claim the needed relation is established simply by a reliable causal process linking first-order mental states with second-order beliefs, and those, like Byrne's, which hold that the subject can move from being in a given first-order state to having an appropriate second-order belief by drawing an inference.

Reflectivists reject all versions of the epistemic approach. This does not mean that they offer no account of how self-knowledge is possible. They offer a different sort of account, one that is primarily metaphysical rather than epistemological. Not just any sort of condition of a person is one whose obtaining involves its being tacitly known to obtain. The reflectivist's task is to explain the nature of various mental states in a way that clarifies why their existence implies that their subject has tacit knowledge of them, and what this tacit knowledge can amount to. I think Moran's idea—that I can have transparent knowledge of my own beliefs because they are expressions of my rational agency—is best understood as offering this sort of account of our doxastic self-knowledge. I do not claim that Moran himself understood his idea in exactly this way, but I have become convinced that his insight takes its best shape in this setting. On this reading, the importance of Moran's work is not primarily that it identifies a connection between doxastic self-knowledge and a special way of *arriving at* beliefs—by 'making up one's mind'—but that it shows how normal doxastic self-knowledge reflects something about what believing *is*.

Briefly, the idea would be that, for a rational creature, believing *P*

[13] The reflectivist can admit that it is possible for me to be in an *alienated* condition in which I hold a belief that *P* which is not consciously accessible, but nevertheless believe, on some other basis, that I believe *P*. Here my self-ascriptive belief is distinct from my first-order belief precisely because it is not grounded in the tacit knowledge that grounds my normal judgements about what I believe.

Proceedings of the Aristotelian Society Supplementary Volume lxxxv
doi: 10.1111/j.1467-8349.2011.00204.x

just is being in a condition of actively *holding P* to be true.[14] The self-determined character of this condition is especially *evident* where a person consciously considers whether *P* and 'makes up his mind', but it is equally real where he holds a belief without conscious reflection. One indication of this is that we expect a rational subject who believes *P* to be able to address the question what convinces him that *P* is true, whether he has consciously deliberated or not. A person will not, of course, always have specific grounds for holding a given belief, but the interesting thing is that, even when he admits to lacking grounds, *he accepts the presupposition of the question*—that he is in a position to speak for whatever grounds he has. Moreover, we do not merely take him to be in a position to give expert *testimony* on this topic. If the grounds he produces are poor, we criticize them, and we address our criticisms to *him*: we ask him why he believes something so outlandish, how he can accept such a manifestly unreasonable argument, etc. We thus normally treat a person's believing *P* on certain grounds (or none) as a circumstance that is in some sense up to him, not merely one about which he is specially knowledgeable. And we do so throughout the duration of his belief, whether or not he has consciously deliberated about it. These are aspects of our privileged relation to our own beliefs about which Byrne's account has strikingly little to say.

What these observations suggest, I think, is that *all* our normal, non-alienated beliefs are, in a perfectly good sense, acts of our capacity to make up our minds.[15] They are all enduring actualizations of our power to evaluate propositions as true, in the light of such grounds as we deem relevant. This evaluation is not an act one performs to *produce* a belief in oneself; it is one's belief itself. And Moran's idea, as I am understanding it, is that this fact about the nature of our believing explains our capacity for transparent knowledge of what we believe. I can normally treat the question whether I believe *P* as tantamount to the question whether *P* because my answer to whether *P* expresses my knowing evaluation of *P* as true (or my refusal so to evaluate it), and unless I am in an alienated condition, my knowingly evaluating *P* as true just is my believing it. Hence my knowing acceptance of *P* constitutes tacit knowledge that

[14] For further discussion, see again Boyle (forthcoming).

[15] Thus I reject Byrne's claim (2011, p. 208) that Moran's conclusion rests on an 'overly restricted diet of examples'.

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

I believe *P*. For suppose I knowingly accept *P*, and understand that what I accept as true just is what I believe. Then I have everything I need to knowingly judge: I believe *P*. All I need do is reflect.

Moran's idea thus supplies the basis for a promising reflectivist account of our transparent knowledge of what we believe. A similar account will be available wherever it is plausible to conceive of a given kind of mental state as constituted by the subject's knowingly evaluating a certain content in a certain way. Thus there appears to be a connection between intending to do *A* and regarding *A* as to be done, desiring some object *O* and regarding one's having *O* as desirable, hoping that *P* and regarding *P* as a possibility whose realization would be good, etc. Even where such attitudes are held without deliberation, a subject's holding them is in general connected with her knowingly accepting (or being inclined to accept) a specific evaluation of a certain matter, and with her being ready to address the question *why* she evaluates the relevant matter as she does. In each case, of course, the connection is complex and in need of further specification. It will be most straightforward where the subject is in a non-alienated condition—one in which she reflectively endorses the relevant evaluation—and this will arguably be the normal case from which others must be understood as departures. My aim here is not to work out the details of a reflectivist account of our knowledge of these attitudes, but simply to indicate the general lines that such an account would follow: it would explain my capacity for transparent knowledge of such states in terms of my ability reflectively to recognize an evaluation I already knowingly make, an evaluation my making which normally just constitutes my being in the relevant state.

On this Moran-inspired reflectivist view, we have privileged self-knowledge of these sorts of attitudes because they are expressions of our capacity for rational self-determination. Their existence normally just consists in our knowingly assessing a certain matter in a certain way, on the basis of whatever grounds we deem relevant; and we have privileged knowledge of them, not because we are in a specially good position to form second-order beliefs that reliably *track* their existence, but because their existence is normally *constituted* by our knowing assessment of the matter in question. There are, however, other kinds of mental states of which we have privileged self-knowledge, but whose existence surely does not consist in our self-determinedly holding an attitude in this way. When it visually appears to

Proceedings of the Aristotelian Society Supplementary Volume lxxxv
doi: 10.1111/j.1467-8349.2011.00204.x

me as if there is a table in front of me, when I feel a throbbing pain in my elbow, or experience the pleasurable rasp and warmth of having downed a shot of bourbon, these are, intuitively speaking, episodes I undergo, not states whose existence consists in my holding a certain self-determined view of some matter. If we are to give a reflectivist account of our knowledge of these states, it must take a different shape. The general reflectivist thesis will still hold: the existence of such states will imply that the subject has tacit knowledge of them. But the account of the nature of the relevant states in virtue of which this connection holds will need to be different.

Thus, if there is a distinction between psychological conditions whose obtaining expresses my capacity for self-determination and conditions whose obtaining expresses my capacity to have my mental state determined through affection, a reflective account of self-knowledge will need to distinguish two kinds of self-knowledge: spontaneous and receptive.[16] This, indeed, is a direct consequence of the distinction between two kinds of psychological conditions, given the reflectivist idea that the very same actualization of my psychological capacities that is my being in state *M* is, under another aspect, my tacitly knowing that I am in *M*. For if being in *M* involves tacitly knowing that one is in *M*, then a basic difference in what grounds the existence of the relevant *M*s will imply a corresponding difference in what makes our privileged self-knowledge of *M*s possible.

I cannot offer a detailed account of receptive self-knowledge here. The basic idea would be that receptivity is simply the capacity knowingly to be affected by an object, and hence that, although I am in the first instance aware of that which is affecting me, my awareness also involves tacit cognizance of my own affection. (This would explain the 'transparency' of these conditions: the fact that I can make a knowledgeable judgement about my own psychological state simply by attending in the right way to something affecting

[16] For further argument in support of such a distinction, see Boyle (2009). Constraints of space prevent me from engaging in detail with Byrne's criticisms of my arguments. Several of them, however, depend on assumptions—about the soundness of the epistemic approach to self-knowledge, and the viability, specifically, of inferentialism—that I have already disputed.

I will add one further remark. Byrne (2011, p. 213) suggests that a view which rejects the Uniformity Assumption, as mine does, should predict the possibility of dissociations between one kind of self-knowledge and another. This is mistaken. To claim that there are *different* kinds of self-knowledge is not necessarily to claim that the relevant kinds are *independent*, and thus potentially dissociable. In Boyle (2009) I argue specifically that our knowledge of ourselves as patients depends on our knowledge of ourselves as agents.

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

me.[17]) Such an account would need to be supplemented by a story about those psychological conditions, like pleasure and pain, which are not presentations of an object affecting me, but modes of awareness of my own condition qua creature-subject-to-affection. It would also need to explain how the capacity that puts me in a position to experience affection by objects also puts me in a position to experience a deficiency in some vitally important mode of affection (as, for example, when I feel hunger). All this lies beyond the scope of the present discussion, but I hope it is plausible that such a project might be carried through.

Let me conclude with a general remark about our topic: self-knowledge. Byrne (2011, p. 201) begins by raising a doubt about the aptness of this term, but our discussion suggests a better motivation for it than the one he considers. The point is not that this is knowledge of a special sort of *object* (a 'self'), but that it is a distinctive sort of *knowledge*—knowledge whose existence is an essential aspect of the condition known, rather than a further state of affairs over and above the obtaining of this condition. There is, I think, a powerful philosophical undertow that pulls us toward conceiving of self-knowledge as not different in principle from knowledge of other objects, however special the epistemological details may be. Epistemic approaches to self-knowledge accept this idea. If reflectivists are right, however, to accept this conception of self-knowledge is already to lose hold of the topic. The proper topic of an account of self-knowledge is a distinctive *kind* of knowing, one that does not track some independent condition, but rather reflects on aspects of the intrinsically known agency and patiency of the knowing subject herself.

[17] Thus the fact that both perceptions and beliefs can be known transparently does not imply that we must give a uniform treatment of these two sorts of self-knowledge. I think, incidentally, that Byrne misrepresents Evans in reading him as proposing a uniform treatment here (cf. Byrne 2011, pp. 212–13). Evans does observe that both sorts of condition can be known transparently, but he goes on to remark that

> [t]he cases are different in that in this [perceptual] case there is something (namely an internal informational state of the subject), distinct from his judgement, to which his judgement aims to be faithful [whereas in the case of knowledge of his own beliefs, there is no such thing]. (Evans 1982, p. 230)

There are aspects of this account that I would want to resist, but the general spirit is congenial to me: Evans thinks that in the perceptual case I seek accurately to describe a state I do not constitute, whereas in the doxastic case there is only the thing I freely constitute, my judgement on *P* itself.

Proceedings of the Aristotelian Society Supplementary Volume lxxxv
doi: 10.1111/j.1467-8349.2011.00204.x

Department of Philosophy
Emerson Hall
Harvard University
Cambridge, MA 02138
USA
boyle2@fas.harvard.edu

References

Augustine 2002: *De Trinitate*. Translated by Stephen McKenna as *On the Trinity, Books* 8–15, ed. Gareth B. Matthews. Cambridge: Cambridge University Press.

Boyle, Matthew 2009: 'Two Kinds of Self-Knowledge'. *Philosophy and Phenomenological Research*, 77:1, pp. 133–64.

——forthcoming: 'Active Belief'. To appear in a special issue of the *Canadian Journal of Philosophy* dedicated to the theme: Belief and Agency.

Byrne, Alex 2005: 'Introspection'. *Philosophical Topics*, 33:1, pp. 79–104.

——2011: 'Transparency, Belief, Intention'. *Proceedings of the Aristotelian Society Supplementary Volume* 85, pp. 201–21.

Dretske, Fred 1994: 'Introspection'. *Proceedings of the Aristotelian Society*, 94, pp. 263–78.

Evans, Gareth 1982: *The Varieties of Reference*, ed. John McDowell. Oxford: Clarendon Press.

Gallois, André 1996: *The World Without, the Mind Within: An Essay on First-Person Authority.* Cambridge: Cambridge University Press.

Harman, Gilbert 1990: 'The Intrinsic Quality of Experience'. *Philosophical Perspectives*, 4, pp. 31–52.

Kant, Immanuel 1997: *Critique of Pure Reason*, ed. and trans. Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.

Locke, John 1975: *An Essay Concerning Human Understanding*, ed. P. H. Nidditch. Oxford: Oxford University Press.

Martin, M. G. F. 1998: 'An Eye Directed Outward'. In Crispin Wright, Barry C. Smith and Cynthia Macdonald (eds.), *Knowing Our Own Minds*, pp. 99–121. Oxford: Oxford University Press.

Moore, G. E. 1903: 'The Refutation of Idealism'. *Mind*, 12:4, pp. 433–53.

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

Rödl, Sebastian 2007: *Self-Consciousness*. Cambridge, MA: Harvard University Press.

Shoemaker, Sydney 1994: 'Self-Knowledge and "Inner Sense". Lecture III: The Phenomenal Character of Experience'. *Philosophy and Phenomeno-*

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x

logical Research, 54:2, pp. 291–314. Reprinted in his *The First-Person Perspective and Other Essays*, pp. 246–68. Cambridge: Cambridge University Press, 1996.

——2009: 'Self-Intimation and Second Order Belief'. *Erkenntnis*, 71, pp. 35–51.

Tye, Michael 2000: *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.

Proceedings of the Aristotelian Society Supplementary Volume LXXXV
doi: 10.1111/j.1467-8349.2011.00204.x