Self-knowledge and the Limits of Transparency

Jonathan Way

Published in *Analysis* 67 (July 2007): 223-30.

It seems that many of our attitudes are *transparent*, in the following sense: we can come to know that we have an attitude *M* by considering a question about the content of *M*. This is clearest in the case of belief, as is illustrated by the following oft-quoted passage of Gareth Evans's,

> ....in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward – upon the world. If someone asks me 'Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' (Evans 1982: 225)

But it is also true, to varying degrees, of other attitudes as well. As Dorit Bar-On points out,

> If asked whether I am hoping or wishing that *p*, whether I prefer *x* to *y*, whether I am angry at or afraid of *z*, and so on, my attention would be directed at *p*, *x* and *y*, *z*, etc. For example, to say how I feel about an upcoming holiday, I would consider whether the holiday is likely to be fun. Asked whether I find my neighbour annoying, I would ponder her actions and render a verdict. (Bar-On 2004: 106)

This remarkable fact – that we appear to be able to answer questions directed at one subject matter by considering questions directed at another – has played a leading role in several recent accounts of self-knowledge. Thus Richard Moran claims (2001: 150) that transparency is 'the fundamental feature of self-knowledge', and argues at length that it is transparency that marks the difference between those attitudes which can be objects of

'ordinary' self-knowledge and those attitudes which can be known, if at all, only through the kind of evidence which is equally available to a third-person.[1] And Moran and others also claim that understanding transparency is the key to understanding those features of ordinary self-knowledge – such as immediacy, authority, and its relation to rationality – which have traditionally seemed problematic to philosophers.[2]

In this paper I raise some problems for these claims. I argue that, on a natural understanding of transparency, there can be attitudes which are knowable in the ordinary way but which are not transparent. If this is right then the attempt to account for immediacy, authority and the rest of it via transparency will not always work. Nor will it be correct to say that transparency distinguishes ordinary and 'merely attributional' self-knowledge.

My strategy is as follows. I begin by giving an interpretation of the transparency claim, which seems to me both to be plausible and to serve the purposes of the theorists who place such weight on it. I then explain how, so interpreted, the claim rests on an assumption about the relationship between a subject's reasons and the attitudes that those reasons permit which is widely accepted to be false of intentions, and is controversial of beliefs. The denial of this assumption allows us to generate examples of rational attitudes, over which we have paradigmatic self-knowledge, which are not transparent. Moreover, even if the relevant assumption about rational belief is insisted upon, the problem for our knowledge of our intentions still has ramifications for the account of our knowledge of

[1] Moran 2001: ch. 3, 4. See also Falvey 2000.
[2] Moran claims (2001: 100–101) that transparency is the 'feature of the first-person position that accounts in part both for the way in which first-person reports are made without appeal to evidence, and why…(self-knowledge)…of this sort should be bound up with the rationality of the person'. See Moran 2003, 2004 for further clarification of this strategy. Other writers who give transparency a starring role in their account of self-knowledge include Gallois (1996) and Fernandez (2003). For discussion see Brueckner 1998 and Zimmerman 2004, respectively.

our beliefs, given the prima facie desirability of having an account of self-knowledge which is at least to some extent unified.

1. The Interpretation of Transparency

So far we have characterized transparency vaguely, in terms of the capacity to answer a question about whether or not one has some particular attitude by consideration of a question about its content. We now need to be a little more specific about what this amounts to. In particular I want to make explicit three assumptions about transparency that the worry I raise will rely on.

First, transparency is supposed to be a way of gaining knowledge about one's attitudes. The idea is that we can make knowledgeable judgments about a state of mind by answering a question about its content. It is hard to see how one could deny this if one sees transparency as being in whatever way central to self-knowledge.

Second, to say that we answer one question *by* answering another is to say that the answer we give to the latter question determines the answer we give to the former. Thus when I ask myself whether I believe that *p*, I answer by answering the question of whether *p* in the sense that my answer to the latter question determines my answer to the former question. If I answer whether *p* in the affirmative then my answer to whether I believe that *p* is also in the affirmative, and if my answer to whether *p* is either *no* or *I don't know*, then my answer to whether I believe that *p* is that I don't.[3]

[3] I make the simplifying assumption that these are the only three answers I could give to the question of whether p, thus ignoring answers like 'most likely, p' and 'possibly p'. Weakening this assumption would only help my argument.

Third, the question of whether I have mental state *M* is in the first instance transparent to the question of whether *to* have *M*. This claim requires some defence – after all, in the central example of transparency a question about one's belief is answered by answering a question about its content, rather than a question about what to believe. But notice that this characterization of transparency does not allow us to extend it to other mental states. After all, it is clearly not the case that one can answer the question of whether one intends *p* (let alone whether one hopes, or fears, that *p*) by answering the question of whether *p*. To mention just one difficulty, it is possible to foresee that one will do something without intending to do it. Thus if one did try to answer questions about one's own intentions in this way, one could easily be led to mistaken conclusions.

But nonetheless I do seem to be able to answer questions about my own intentions by looking 'outward', in a way analogous to how I answer questions about my own beliefs. When I ask myself how I intend to vote in the forthcoming elections I reflect on the relevant candidates, and their parties and policies, and I can thereby come to a judgment about what I intend to do. I do not make an 'inward glance', or consider evidence about my own behaviour. Plausibly, I here answer the question about my intention by answering the question of what to intend.

And we can understand the transparency of belief in much the same way. After all, when, in Evans's example, I consider the 'phenomena…I would attend to if I were answering the question "Will there be a third world war?"' I am considering *reasons* to think that there will be a third world war. Thus I am – perhaps only implicitly – considering the question of whether to believe that p. It is just that this question, in turn, is answered by considering the question of whether p. Thus both belief and intention can be seen as, in

the first instance, transparent to a normative question about whether to have the attitude in question.[4]

2. Transparency, Akrasia and Uniqueness

The understanding of transparency given above makes it reasonably clear how it is that we could come to know that we have some mental state by considering the content of that state. The basic procedure is that we ask ourselves whether to have attitude *M*, and if we conclude that *M* is the attitude to have, we judge that we have it, and if we conclude that it's not, we judge that we don't.

There is an obvious objection to this idea: agents don't always have the attitudes that they judge are the attitudes to have. I can intend to have another drink, while knowing full well that this is not the thing to be doing. Thus it seems that coming to judgments about our attitudes by reflection on what attitudes to have will lead us astray in such cases. And the possibility of such attitudes places Moran's claim that transparency is central to self-knowledge, and the key to understanding immediacy and authority, in doubt. For after all, can I not know – in the ordinary, first-person way – that I intend to have another drink, even though this attitude is not transparent?

Moran's response to this objection is to deny that we 'speak with first-person authority' about our akratic attitudes (2001: 128).[5] An exploration of his discussion and defence of

[4] This is certainly the way Moran understands it: 'transparency requires…the deferral of the theoretical question "What do I believe?" to the deliberative question "What am I to believe?"' (2001: 63). See also his discussion at 2003: 405. Gallois' understanding of the matter is slightly different but he does claim (1996: 140–41) that 'giving an affirmative answer to ["is there, on balance, reason to believe p?"] enables one to give an affirmative answer to ["is p true?"]…[which]…enables one to give an affirmative answer to ["do I believe p?"]'. This claim, and its analogues for the other attitudes plays a crucial role in his account. See 1996: chap. 7.

[5] Gallois (1996: 147) takes a similar line. Fernandez, who only considers beliefs, denies (2003: 368) that epistemic akrasia is possible. As Zimmerman 2004 shows, this is problematic.

this claim would lead us too far astray and, I shall argue, would anyway be redundant.[6] This is because there are *rational* attitudes – which can be recognized as such by their agents - which are just as resistant to transparency, and yet which can uncontroversially be the objects of ordinary self-knowledge. By focussing on these examples we can generate just the same problems for Moran, while sparing ourselves the need to resolve the considerable complexities involved in trying to understand akrasia. I can start to explain this by pointing out a crucial assumption of the transparency procedure.

Suppose I ask myself whether to have some attitude *M* and I judge that, although *M* is permissible it is not required – what is required is only that I have *M* or that I have *N*. In such a case there is no attitude which is *the* attitude to have, and thus it is hard to see how consideration of the questions of whether to have *N*, and whether to have *M*, can lead me to a judgment about which I do have. My consideration of these questions is inconclusive in the relevant respect. The point can be reinforced: if, in such a case, consideration of the question whether to have *M* could lead me to judge that I have *M*, then consideration of the question whether to have *N* could also lead me to judge that I have *N*. But it might well be the case that although I ought to have *M* or *N*, I ought *not* to have both. Consideration of the question of whether to have *M* cannot then be enough to allow me to come to know that I have *M*. This suggests that transparency will only work in general on the assumption of *uniqueness*:

> Given a set of reasons *R*, bearing on S's having some attitude *M*, either S ought to have *M*, or S ought not to have *M*.[7]

---

[6] For the record, and despite Moran's interesting and insightful discussion, I remain unconvinced. See Owens 2003 for further discussion of this problem.

[7] Some of Moran's claims about transparency seem to reflect this: 'In deciding what to do, [a person's] gaze is directed "outward", on the considerations in favour of some course of action, *on what he has most reason to do*' (2001: 105, my italics),

But it is widely agreed that uniqueness does not hold for intentions. For example, as Michael Bratman has emphasized, it is often the case that, like Buridan's ass, we need to choose between two or more options which are equal in respect of the reasons favoring them. In Bratman's example (1987: 23), he must choose between taking route 101 and route 208 to San Francisco, where the two routes are equally attractive, and equally efficient as a means to his end. Supposing that going to San Francisco is what he ought to be doing, we can conclude that he ought either to take route 101 or route 208, but not that he ought to take the 101, nor that he ought to take the 208, nor that he ought to take both. Thus we can generate counter-examples to transparency.

Suppose that Bratman decides to take the 101, and thus that he intends to do so. This intention may be one over which he exhibits paradigmatic self-knowledge – his knowledge of his intention will be both immediate and authoritative – but it will not be transparent. He cannot answer the question of whether he intends to take the 101 by considering the question of whether to take the 101. As he recognizes, the reasons bearing on the question of whether to take the 101 do not determine that taking the 101 is the thing to do – he could just as well take the 208. Thus any answer he gives to the question of whether to take the 208 he can equally give to the question of whether to take the 101. So he cannot know whether he intends to take the 101 by considering the question of whether to take the 101.

Cases with this kind of structure are pervasive in the practical realm.[8] The example of choices between equally good options is just one kind of example. Other clear examples are cases of choice between incommensurable options, and choice when one is ignorant

[8] As Joseph Raz (1999: 100) puts it, 'most of the time people have a variety of options such that it would accord with reason for them to choose any one of them and it would not be against reason to avoid any of them'.

or uncertain of what can be said in favour of one's options. More controversial examples include: deciding to do something 'for no reason'; deciding to φ on the grounds that φing is in the relevant respect *good enough*, even though ψing would be in that respect better; deciding what to do in a case where one of one's options is supererogatory. In each of these cases an agent will be able to form a rational intention, over which he exhibits paradigmatic self-knowledge, but, as he will not affirm that his intention is the attitude to have, it will not be transparent.

Cases with this structure are more difficult to generate in the theoretical realm, and thus uniqueness is more controversial for belief than intention. Consider what seems to be the most straightforward parallel to Bratman's case, a case where one has equal reason to believe two inconsistent propositions. For example, I have equal reason to believe that there are even number of students on campus right now, and that there are an odd number of students on campus right now. Yet, in contrast to the intention case, it is not rational for me to believe either of these propositions. The rational thing to do is to suspend judgment. Thus I can answer the question of whether I believe that there are an odd number of students on campus right now – I consider whether or not there are, realize that I ought not to believe that there are, and thus judge that I don't believe that there are. So we have no quick and easy counter-example to transparency.

Nonetheless the claim that there is always a uniquely correct attitude to take towards p, when one is considering whether p, remains a strikingly strong claim. It is far from obvious why there cannot be evidential states good enough to permit believe in p, without requiring belief in p, and thereby also permitting suspension of belief in p. And indeed, many philosophical theories of rational or justified beliefs make room for this. Thus, according to coherentists, a belief is justified just in case it is part of a network of mutually coherent beliefs. But as coherence may be achieved in different ways, this

leaves open the possibility that both belief in p, and suspension of judgment in p, will make for similar degrees of coherence. And according to proponents of conservatism, a belief is justified unless one has reason to abandon it. Again, this allows that both belief in p, and suspension of belief in p, may be acceptable given one's evidence.[9]

Thus it seems that the defender of the transparency-based approach to self-knowledge is committed to a thesis which seems clearly false of intentions, and is controversial for belief. Moreover even if he is willing to accept, and defend, uniqueness for beliefs, it is hard to see how he could do this for intentions. Thus he has real difficulties in generalizing his account beyond (one aspect) of the cognitive realm.

3. Another Interpretation of Transparency?

Faced with this objection a defender of the centrality of transparency may well be tempted to reject my interpretation of that claim. But it is hard to see how else to understand transparency, if it is to yield the result that answering a content-directed question is a way of coming to know one's own mind. As the kind of cases discussed above show, for example, it would be no good to say that attitudes need only be transparent to the question of whether there is reason to have them, or even sufficient reason to have them.[10] If, as seems likely, there are cases where there is sufficient reason to adopt more than one attitude, such a procedure will be prone to lead to mistaken conclusions.

---

[9] See White 2005 for further detail on these points and further examples of philosophical accounts of rational belief which deny uniqueness. White also offers a number of interesting arguments in defence of uniqueness for beliefs.

[10] Moran does sometimes suggest claims like this. For example, he says (2001: 105) that the question of what to do is 'answered by the 'outward-looking' consideration of what is good, desirable, or feasible to do'. Much of what Gallois says also suggests this interpretation. See, for example, Gallois 1996: 139–147.

Thus, pending further clarification, we are left without a tenable account of transparency which vindicates the claims that Moran and others have made for it.[11]

References

Bar-On, D. 2004. *Speaking My Mind*. Oxford: Clarendon Press.

Bratman, M. 1987. *Intentions, Plans and Practical Reason*. Cambridge, Mass.: Harvard University Press.

Brueckner, A. 1998. Moore Inferences. *Philosophical Quarterly* 48: 366–69.

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

Falvey, K. 2000. The Basis of First-Person Authority. *Philosophical Topics* 28: 69–99.

Fernandez, J. 2003. Privileged Access Naturalized. *Philosophical Quarterly* 53: 352–72.

Gallois, A. 1996. *The World Without, the Mind Within*. Cambridge: Cambridge University Press.

Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Cambridge, Mass.: Harvard University Press.

Moran, R. 2003. Responses to O'Brien and Shoemaker. *European Journal of Philosophy* 11: 402–19.

Moran, R. 2004. Replies to Heal, Reginster, Wilson, and Lear. *Philosophy and Phenomenological Research* 69: 455–72.

Owens, D. 2003. Knowing Your Own Mind. *Dialogue* 42: 791–98.

Raz, J. 1999. *Engaging Reason*. Oxford: Oxford University Press.

White, R. 2005. Epistemic Permissiveness. *Philosophical Perspectives* 19: 445–59.

Zimmerman, A. 2004. Unnatural Access. *Philosophical Quarterly* 54: 435–38.

[11] I would like to thank Matthew Hanser, Michael Rescorla, Aaron Zimmerman and especially Kevin Falvey, for comments on various versions of this paper and discussion of related issues.