

## The Silence of Self-Knowledge

Johannes Roessler

[Forthcoming in *Philosophical Explorations*]

### Abstract

Gareth Evans famously affirmed an explanatory connection between answering the question whether *p* and knowing whether one believes that *p*. This is commonly interpreted in terms of the idea that judging that *p* constitutes an adequate basis for the belief that one believes that *p*. This paper formulates and defends an alternative, more modest interpretation, which develops from the suggestion that one can know that one believes that *p* *in* judging that *p*.

Keywords: self-knowledge; rationality; belief; transparency; cognitive achievement.

According to some philosophers, it's a mark of psychological self-knowledge that the subject lacks any awareness of its source. Self-knowledge is acquired imperceptibly or, in Brian O'Shaughnessy's phrase, 'silently'. (2000, 108) The first-person perspective affords no insight into how we know what we believe, desire or experience. Other philosophers disagree. They insist that reflection on what we *do* when we reflect on our own current mental states reveals that we rely on an epistemic basis in terms of which our knowledge is readily intelligible (though of course the finer points of an explanation of self-knowledge will be reserved to philosophers).

An influential example of such 'second-order reflection' is a brief passage from Evans's *Varieties of Reference*, which may be encapsulated in the thesis that answering the question whether *p* can be a way of (or a 'procedure for') gaining knowledge of whether one believes that *p*. (1982, 225) I'll call this Evans's thesis (ET). ET has been glossed and elaborated in a variety of ways. Richard Moran takes Evans's point to be that 'a first-person present-tense question about one's belief is answered by reference to (or consideration of) the same reasons that would justify an

answer to the corresponding question about the world.’ (2001, 62) Bernard Williams has argued that ‘I am confronted with my belief as what I would spontaneously assert.’ (2003, 76) According to Alex Byrne , ‘I infer that I believe that there will be a third world war from the single premise that there will be one.’ (2011, 203) There are significant differences between these theories, but there is also a common theme, the existence of an explanatory link of some kind between judging whether p and possession of knowledge whether one believes that p. Moreover, given the way Evans and his followers arrive at ET, namely by reflection on what we — not just philosophers but reflective thinkers in general — intentionally do when we answer questions about our own current beliefs, it would be surprising if a subject whose self-knowledge was based on this method were wholly unaware of how she acquired it. This, in turn, would make it puzzling how anyone, let alone a philosopher with peerless powers of self-scrutiny such as Brian O’Shaughnessy, could deny that there is any ‘cognitive path via which this knowledge is reached’ (2000, 105).

My aim in this paper is to dispel this puzzle by arguing for a middle course between the two views on the intelligibility of self-knowledge. Philosophers tend to be interested in making our possession of knowledge in a given area intelligible in a distinctive way: by appeal to a basis such that there is an intelligible link between a belief’s being held on that basis and the belief’s ‘status as knowledge’.

O’Shaughnessy was right, I will suggest, that there can be no epistemic basis in that sense, open to reflection, for knowledge of one’s current beliefs. The problem with the idea of such a basis, I will argue, is that one could not self-consciously draw on it, or find one’s knowledge intelligible in terms of it, without irrationality. This, though, should not lead us to reject ET. For while ET is standardly interpreted as the idea that

judging that *p* provides one with an epistemic basis for a self-ascription of the belief that *p*, such an interpretation is by no means mandatory. On what I will call a modest construal of ET, someone's answering the question whether *p* — say, by judging that *p* —, can make her possession of self-knowledge intelligible, given that one is usually aware of believing that *p* *in* judging (or sincerely asserting) that *p*. In the first three sections I develop the modest reading. In the second half of the paper I present my challenge to the standard ambitious construal of ET.

### **1. Two readings of Evans's Thesis**

That ET makes some kind of explanatory claim seems clear enough. If you gain something by using a certain 'way of gaining' things of the relevant kind, it should certainly be possible to explain your possession of the thing by appeal to how you got it. The question is how the explanation actually works. One way to understand ET would be to think of it as an alternative answer to the very question addressed by a 'perceptual model' of self-knowledge. The rejection of such a model is of course a central theme of Evans's discussion of 'mental self-ascriptions'. What Evans rejects, you might say, is the model's misguided *answer* to a question that is a perfectly good question to ask — the correct answer to which is provided by ET. The natural way to put the question is: How do we know what we currently believe?

Byrne's account provides a clear example of this kind of reading. According to Byrne, using Evans's 'procedure' involves 'inferring' that one believes that *p* from a single premise, viz. '*p*'. Following Gallois (1996), he calls this putative form of inference the 'doxastic schema'. His project is to offer an account of how it is possible for reasoning in accord with the schema to yield knowledge. The chief difficulty he sees

is that the schema is ‘neither deductively valid nor inductively strong’. (2011, 204) Byrne’s suggestion, briefly, is that the schema can be seen to be ‘strongly self-verifying’, given that ‘inference from a premise entails belief in that premise’. (2011, 206) This in turn is held to provide grounds for thinking that reasoning in accord with the schema — self-ascribing the belief that p on the basis of the premise that p — is ‘knowledge-conducive’.

A salient feature of this kind of explanation is the central role it assigns to the subject’s belief that she believes that p. What explains her knowing that she believes that p, on this kind of account, is that she comes to *believe* that she believes that p on a certain basis, the use of which can be seen to be ‘knowledge-conducive’. Thus the explanation turns on an analysis of the distinctive epistemic warrant or entitlement conferred on a self-ascription of belief by fact that it is ‘based’ on the judgement that p. An important implication of this is that the explanation will not only shed light on the etiology of A’s knowledge (she came to know that she believes that p by answering the question whether p) but will also give us a substantive, theoretically informed reason to think that what A has is indeed *knowledge*. This is part of what makes it natural to think of the account as a good answer to the question ‘how does A know she believes that p?’ For, as Austin pointed out, it’s a characteristic feature of the question ‘how does A know?’ that it can be asked as a ‘pointed’ question, bringing into play the possibility that A may not actually be in a position to know that p, but may just believe it, possibly on the basis of rumour or faulty reasoning. (1961,

78) Byrne's account, if successful, would certainly deliver a reassuring answer to Austin's pointed question.<sup>1</sup>

I want to suggest that ET does not commit one to the availability of an explanation of this kind, in terms of an epistemic basis for self-knowledge. There is an alternative, more modest way to fill out ET. I should say that I'm not claiming that the modest reading is what Evans actually had in mind. (As far as I can see, the question is moot.) My immediate point is just that it is not obvious that the way in which Evans's 'procedure' helps to make knowledge of one's beliefs intelligible has anything to do with an explanation of the 'epistemic status' of second-order beliefs in terms of their epistemic basis. I labour this point in this and the next two sections. In sections 4-6 I present grounds for scepticism about ambitious versions of ET.

Suppose A knows that she believes that p. The thrust of Evans's discussion is that this fact may be unsurprising in the light of another fact, that A has just answered the question whether p. Note that what primarily explains A's self-knowledge, in this schema, is not his judging that p but his answering the question *whether p*. We should be careful to distinguish the two things. Answering a question may take time. It may, as Evans emphasized, involve many kinds of mental activity, such as observation, deliberation or recollection. You may be *in the process* of answering the question whether p without ever reaching a verdict, due to being interrupted. On the other hand, it is hard to think of what might be involved in being interrupted in judging that

---

<sup>1</sup> For further examples of the ambitious reading of ET, see Peacocke (1996), Moran 2001 (support for this interpretation will be offered in section 3 below), Boyle (2009). For illuminating discussion of Peacocke's version, see Martin (1996).

p. The latter is naturally thought of as an event — an event that may constitute the culmination of the process of answering the question whether p.

What I call a modest construal of ET is the suggestion that A knows that she believes that p not on the basis of judging that p but *in* judging that p. There may be more than one way to develop this idea, but the following seems to me a natural and attractive elaboration. In doing something intentionally one generally knows what one is doing, under descriptions corresponding to the contents of at least some of one's intentions. Now sincerely asserting that p is an act informed (partly) by the intention to express one's belief that p. Hence the subject will typically be aware of expressing her belief that p. The point can be extended to the case of judgements that involve no overt utterance. Judging that p, it is plausible to suppose, involves saying that p in 'inner speech' (if not in outer speech). Perhaps what this means is that one *imagines* saying that p. But whatever the correct analysis, saying that p in 'inner speech' is often informed by the intention to express one's view that p, even if only for one's own benefit. Indeed acts of saying that p in 'inner speech' that are not so informed would arguably not be cases of *judging* that p (instead they may, say, be part of imagining a conversation). Quite generally, then, judging that p involves being aware of expressing one's conviction that p. Of course this schematic point conceals many complexities. For one thing, whether a particular thought or assertion amounts to a judgement, or perhaps merely expresses an assumption whose merits one is exploring, may not always be entirely clear to one at the time.

For the moment, let's set such complications to one side. The important point, for now, is that there is a natural way to understand ET that involves no reference to any

putative epistemic basis for self-knowledge. For it is at least not obvious — and indeed there is a strong philosophical tradition that rejects the idea — that knowledge of what one is intentionally doing is to be explained in terms of any epistemic basis. On the current construal, the sense in which answering the question whether *p* may be a ‘way of gaining knowledge’ of one’s belief that *p* can be spelled out as follows: answering the question whether *p* is a process that, if one does believe that *p*, and is not interrupted, can be expected to issue in one’s judging that *p*; and in doing so, one will be aware of expressing one’s view that *p*, hence aware of believing that *p*.<sup>2</sup> (Of course the ‘way of gaining knowledge’ will also be available in the special case where answering the question whether *p* leads one to *form* the belief that *p*.)

It is important to distinguish two kinds of intentions at work in this account. There is the intention to answer the question whether *p*. And there is the intention to express one’s view that *p*. The former will of course be present from the moment the process of answering the question whether *p* gets going. The latter, though, may not. The formation of the intention to assert *that p* (in outer or inner speech) may be more or less concurrent with the event of asserting that *p*. Compare the case of trying to recall Hume’s date of birth: you may no sooner acquire the intention to say ‘1711’ than you blurt it out. Borrowing a helpful distinction from Bernard Williams, we might say that an assertion can be spontaneous ‘as to what’, even if it is not spontaneous ‘as to

---

<sup>2</sup> There are two ways to think about the relation between the two kinds of awareness. It might be said that on the current account, knowledge of what one believes is based on a valid deductive argument (one infers that one believes that *p* from the premise that one is expressing one’s belief that *p*). Alternatively, we might think of one’s awareness of expressing one’s belief that *p* as amounting to a way of being aware of believing that *p* (which, it might be suggested, is naturally expressed by the idiomatic use of ‘I think that *p*’, with the implication that one believes that *p*.)

whether'.<sup>3</sup> Your saying '1711', for example, may be premeditated 'as to whether' — it may be the realization of a prior intention to say when Hume was born — , yet spontaneous 'as to what': you may not have a prior but only a proximal intention to say '1711'. This points to a basic sense in which one may, by the use of Evans's 'procedure', *learn* what one believes. As Williams puts it, 'in the simplest case I am confronted with my belief as what I would spontaneously assert' (2002, 76). The important point is that the 'confrontation' here need not be a matter of being presented with something on the basis of which one comes to know what one believes.<sup>4</sup> It may just be a matter of intentionally (hence knowingly) expressing one's belief that p.<sup>5</sup>

A full defence of ET, as construed on the modest reading, would need to take on each of two opposed lines of response. One response insists that no progress whatsoever has been made in rendering knowledge of one's beliefs intelligible so long as we lack an answer to the question of how one knows what one believes, where this would require a philosophical account of the entitlement-conferring basis on which we come

---

<sup>3</sup> Williams' suggestion, more precisely, is that expressions of belief, 'in the most basic case', are 'involuntary as to what', though not 'involuntary as to whether'. (2002, 75) It is not clear from Williams's discussion whether he takes 'involuntary' to imply 'not intentional'. An alternative, and more attractive, reading, is seems to me, might be that an expression of belief can be involuntary in virtue of being 'spontaneous', where 'spontaneous' contrasts with 'pre-meditated' but not with 'intentional'.

<sup>4</sup> This is one reason it would be a non sequitur to move from the idea that one may gain knowledge of one's attitudes by expressing them to an assimilation of the first- to the third-person perspective. The non sequitur is at least encouraged in the following well-known passage of Dennett's: 'We often do discover what we think (and hence what we mean) by reflecting on what we find ourselves saying — and not correcting. So we are, at least on these occasions, in the same boat as our external critics and interpreters, encountering a bit of text and putting the best reading on it that we can find.' (1991, 245)

<sup>5</sup> Williams himself does not elaborate on the nature of the 'confrontation': his discussion resembles Evans's in that it does not take a stand on whether self-knowledge is to be rendered intelligible in terms of an epistemic basis.



to form second-order *beliefs*. (On the modest reading, ET is silent on this issue.) The other response questions whether reflection on how we acquire self-knowledge serves any useful philosophical purpose. The difference between perceptual knowledge and self-knowledge, it might be said, is precisely that our possession of the latter does not stand in need of being made intelligible or unsurprising by reference to the way we acquire it; what would be surprising is if someone who believed that *p* did *not* know that she did. Each of these responses deserves a much more detailed discussion than will be possible here. But I want to spell out, and motivate, the idea of a middle course between them, first by considering its bearing on the explanation of impairments of self-knowledge (section 2), and then by examining what is sometimes seen as the central puzzle raised by ET (section 3).

## **2. A cognitive achievement**

One good reason to seek an explanation of self-knowledge is an interest in understanding its limitations. The second line of response just mentioned is surely right that we find the absence of self-knowledge, in respect of one's propositional attitudes, more puzzling than its presence. But it may reasonably be argued that only by understanding the way self-knowledge is ordinarily acquired can we hope to make sense of impairments of the capacity for knowing one's attitudes.

An influential formulation of this line of argument is due to Paul Boghossian. Our conception of our own current mental states, Boghossian argues, is not immune to either ignorance or error. *A* may be unaware of her believing that *p*; again she may falsely take herself to believe that *p*. Evidently in such cases *A* is not in an 'epistemically favorable' position with respect to her current beliefs. In other words,

she is unable to make proper use of a way we ordinarily have of gaining knowledge of what we believe. Boghossian concludes that 'it is only if we understand self-knowledge to be a cognitive achievement that we have any prospect of explaining its admitted shortcomings.' (2008, 154)

One possible concern about this line of thought has to do with Boghossian's conception of the shortcomings of self-knowledge. You might insist that it is part of the nature of 'first-person authority' that there is no such thing as being simply mistaken or ignorant about one's current mental.<sup>6</sup> Still, I think as long as we accept, as we surely have to, that self-knowledge does have shortcomings, even if their nature is very different from, say, the shortcomings of perceptual knowledge, Boghossian's point has a great deal of force.

What I think is questionable is a certain view Boghossian assumes of what it means to understand self-knowledge as a 'cognitive achievement'. For him, the idea that self-knowledge reflects a cognitive achievement is tantamount to the claim that it reflects the subject's 'being in a position to *garner the relevant evidence*' (2008, 154, my emphasis). I want to suggest that there is an unargued assumption here: that the way in which self-knowledge must be intelligible if we are to make sense of its shortcomings has to involve reference to 'evidence'; that is, to some kind of epistemic basis that helps to explain how second-order beliefs secure the 'status of knowledge'. (This will be so whatever substantive conception of evidence Boghossian

---

<sup>6</sup> For example, Jane Heal maintains that even when 'a sincere avowal' is rejected on the basis of convincing contrary evidence, it 'still retains a kind of shadowy credence. We are inclined to say that the person 'sort of' believes or intends as she insists she does.' (2003, 273)

favours.) Another way to put the assumption is to say that epistemology, as traditionally conceived, and disciplines interested in the explanation of failures of self-knowledge share the same basic explanatory project.

This is surely not obvious, and on reflection, I think, not particularly plausible.

Failures of self-knowledge are an important concern of commonsense psychology, as well as disciplines firmly rooted in commonsense psychology, such as psychiatry or psychoanalysis. It is not clear that practitioners of these disciplines do, or should, concern themselves with the epistemic basis of which we might be said to form second-order beliefs. Part of the attraction of the modest construal of ET, it seems to me, is that it provides a more plausible way to think about the explanatory project in which these disciplines are engaged. Put briefly, the project is not concerned with a putative epistemic basis for self-knowledge but with the role of certain 'disabling conditions' that can impede one's ability to express, and hence to know, one's attitudes.

An eloquent and illuminating sketch of this sort of project is due the Third Earl of Shaftesbury:

One would think there was nothing easier for us than to know our own minds and understand what our main scope was, what we plainly drove at and what we proposed to ourselves, as our end, in every occurrence of our lives. But our thoughts have generally such an obscure implicit language that it is the hardest thing in the world to make them speak out distinctly. For this reason, the right method is to give them voice and accent. And this, in our default, is what the

moralists or philosophers endeavour to do, to our hand, when, as is usual, they hold us out a kind of *vocal* looking-glass, draw sound out of our breast and instruct us to personate ourselves in the plainest manner. *Those thoughts he murmurs implicitly to himself and under his breath: 'Oh! a funeral for my uncle would be simply splendid!'* (1999 [1711], 77-8)

The central message of this passage is that self-knowledge can be a cognitive achievement. To know one's attitudes one has to be able to make them speak out, in conversation or conscious thought. Doing so can take an effort, and it may require help from others. Shaftesbury highlights motivation as a possible impediment, but there may be other kinds of conditions that can make it hard to articulate one's attitudes, or may give rise to inauthentic expressions of attitudes.<sup>7</sup> This is not the place to develop or defend Shaftesbury's account. My main point here is that Shaftesbury's is a perfectly intuitive notion of a 'cognitive achievement', and that it does not involve any overt reference to an epistemic basis for self-knowledge. The subject's achievement is to express and thus to know her attitudes. Boghossian is exactly right that this account has implications for the explanation of our ordinary possession of self-knowledge. Motivation, or other disabling conditions, can obstruct the acquisition of self-knowledge by affecting the way such knowledge is normally obtained. But there is no reason to assume that this must involve blocking from view any kind of

---

<sup>7</sup> One such condition might be said to be the subject's state of mind. Compare Peter Hobson's account of deficits in insight in patients with borderline personality disorder: such patients show abnormal 'forms of relating [to other people] that amount to different states of mind' (2002, 169). Their abnormal states of mind have an 'impact on the ability to think' (174), in particular to think about certain emotionally charged issues. A 'restriction in the capacity to think' in turn gives rise to 'a corresponding loss of insight into the sources of one's behaviour.' (175) For suggestive discussion of the role of social interaction — specifically, of ways in which one's conception of one's own attitudes may depend on one's audience — , see Williams (2002), ch. 8.

‘evidence’ we usually rely on in gaining self-knowledge. It may just be that usually we know our attitudes *in* expressing them.

### **3. The problem of the two subject matters**

The argument of the previous section does not, it is important to emphasize, amount to an objection to ambitious construals of ET. I have argued that relative to the explanatory purposes of commonsense psychology, there is no need to think of Evans’s procedure as involving some epistemic basis that explains how second-order beliefs secure the ‘status of knowledge’. It would be entirely consistent with this point to argue that ultimately it is hard to make sense of the idea of a ‘way of knowing’ one’s beliefs that is not underpinned by some kind of epistemic basis. It is sometimes said that ET presents us with a puzzle or paradox: it suggests one can gain knowledge of one subject matter (one’s view about whether p) by reflecting on a completely different subject matter (the question whether p). And it might be argued that a resolution to the puzzle requires us to adopt an ambitious construal of ET; that it turns on an analysis of the way in which addressing the question whether p provides one with an adequate basis for self-ascribing the belief that p.

This line of thought, I think, plays a leading role in Richard Moran’s influential discussion of self-knowledge in *Authority and Estrangement*. Here is Moran’s statement of what he takes to be the puzzle:

The immediate problem with the Transparency claim [i.e. ET] is that it seems to say that I can answer a question about the beliefs of a particular person [viz. my belief on the likelihood of a third world war], which would seem to be an

empirical psychological question, not by considering the evidence about that person, but by considering the evidence about some wholly distinct matter of fact, (..), evidence about the increasing militarization of world politics. This is a real problem. (2003: 405)<sup>8</sup>

Moran's well-known account of the connection between self-knowledge and what he calls the 'deliberative stance' is intended, in part, to solve this puzzle. In crudest outline, Moran's solution runs like this. The 'right sort of link between the two questions' is provided by the assumption that '*what my belief is here*' is '*determined by the conclusion of my reflection on those reasons*' [i.e. the reasons for my first-order belief]. (2003: 405, my emphasis). That assumption would give me 'the right to assume that my reflection on the reasons in favour of rain provided an answer to the question of what my belief about the rain is' (ibd..) A further twist is introduced by the following sceptical worry. 'The conclusion of my reflection' is an instance of what Moran calls my 'explicit thinking'. Now a sceptic might question my right to move from my '*explicit thinking*' about whether p to a claim as to what I *believe* about p. As Moran, adapting Matthew Arnold, puts it: 'the current of his true beliefs and feelings' might 'run its own course and have nothing to do with his explicit thinking'. (2001, 123)<sup>9</sup> But Moran insists the sceptical challenge can be met by reflection on the demands of self-awareness. He argues that self-awareness essentially

---

<sup>8</sup> In *Authority and Estrangement* the puzzle is often expressed by asking how it 'makes sense' for me to answer a question as to what I believe by reflecting the object of my belief. (2000, e.g. 66-67) Following Moran, Boyle refers to Evans's procedure as 'seemingly paradoxical' (2009, 137)

<sup>9</sup> The quote from Arnold echoed in this passage is given in Moran 2005, 356-6, n.7: 'Below the surface-stream, shallow and light, Of what we *say* we feel — below the stream, As light, of what we *think* we feel — there flows With noiseless current strong, obscure and deep, The central stream of what we feel indeed.'

involves thinking of oneself as a deliberator, which in turn commits one to taking one's deliberation to be effective in determining one's attitudes. (See 2001, ch. 4; 2003)

Moran's is sometimes interpreted as offering a 'non-epistemic' explanation of self-knowledge. (Reed 2010) This, it seems to me, is misleading. The stated aim of Moran's theory is to 'explain and vindicate avowal as a privileged form of knowledge of oneself.' (200, 134) More specifically, the idea of an entitlement to form second-order beliefs *on the basis* of one's 'explicit thinking' is absolutely central to the theory. Moran would presumably reject Byrne's description of the transition as an 'inference', but there can be no doubt that his theory is committed to the existence of some such transition: it is on the basis of reflection on our first-order reasons that he thinks we self-ascribe beliefs, and philosophical reflection on the nature of the basis reveals that we are entitled to do so. In the terms introduced earlier, Moran's is an ambitious construal of ET.

Later I will raise some doubts regarding Moran's conception of the putative basis for self-ascriptions (section 6). My immediate question is what we should make of the problem of the 'two subject matters'. Is an ambitious construal of ET forced upon us by that problem, as Moran's argument may seem to suggest? The right response to that question, I think, is to take a closer look at the problem itself. On reflection, it seems to me, it is less obvious than Moran assumes that the 'two subject matters' present us with a real puzzle at all. The question we should ask is: relative to what background assumption does it look baffling that you may answer an empirical question not by reflection on evidence relevant to it, but by reflection on evidence

relevant to some ‘wholly distinct matter of fact’? Actually, of course, it is not self-evident (and it would be question-begging for Moran’s argument to assume from the outset) that using Evans’s procedure *essentially* involves reflection on, or even just possession of, relevant first-order reasons. But even if we consider a case, such as Evans’ third world war example, where such reflection undoubtedly occurs, it is by no means obvious what makes it puzzling. Moran asks ‘how it makes sense’ for me to answer the question of what I believe in this way. At the risk of sounding facetious, I think it is worth bearing in mind that ordinary members of the public are not troubled by that problem. If you ask a political pundit for her views on the imminence of a third world war, you would be reluctant to follow this up with the question: ‘but how can it make sense for you to answer that question on the basis of reflection on the wholly unrelated subject of international relations?’ If there is a (philosophical) problem here, it is presumably generated by something like the following assumption:

(1) Empirical questions should be answered on the basis of relevant evidence.

Then one reaction to the puzzle would be to suggest that the case of answering questions about one’s beliefs shows that (1) cannot be sustained as a matter of complete generality. This would be to repulse rather than to resolve the puzzle. The claim would be that we should reject the assumption that can make ET *look* puzzling. Now if answering a question on the basis of evidence involves inference (as a naïve reading would suggest), then Moran, I take it, would actually agree with the repulsing response. His insistence that the ‘two subject matters’ represent a ‘real problem’, I think, reflects a commitment to the following variant of (1):



(2) Empirical questions should be answered by drawing on some relevant epistemic basis.

The puzzle generated by (2) is the ‘real problem’ that Moran’s theory aims to resolve. But note that the claim that there *is* a real problem here simply assumes that (2) holds with complete generality. In the absence of any support for (2) — and it is hard to find any such support in Moran’s discussion<sup>10</sup> — a repulsing response to Moran’s puzzle would be a tempting option. What emerges from this analysis is that the alleged problem of the ‘two subject matters’ can give us no reason to pursue an ambitious construal of ET. For the whole problem is *premised* on the assumption that knowledgeable self-ascriptions of belief must be intelligible and defensible in terms of an epistemic basis.

You might say that Moran’s sceptical scenario does lend support to that assumption. If it is a standing possibility that one’s attitudes may ‘run their own course’, a course not in line with one’s ‘explicit thinking’, does this not show that an episode of ‘explicit thinking’ can never simply *involve* an awareness of one’s attitudes but can be, at best, a basis for self-ascriptions of attitudes? For reasons familiar from debates about perceptual experience, I think this line of thought involves a non sequitur. (See McDowell 1982.) The possibility of ‘inauthentic’ episodes of thinking does not imply that the content of one’s awareness of what one is thinking must be neutral on the

---

<sup>10</sup> One might expect to find such support in the first chapter of *Authority and Estrangement*, where Moran defends a view of self-knowledge as representing a ‘cognitive achievement’, in opposition to ‘deflationary’ approaches to self-knowledge. But note that this discussion is predicated on the assumption I have been questioning in the previous section, that the only alternative to an explanation of self-knowledge in terms of some epistemic basis for second-order beliefs consists in the deflationary repudiation of the notion that self-knowledge involves a ‘cognitive achievement’.

authenticity or otherwise of the episode of thinking. Instead we might say that *in general* such awareness involves an awareness of having the attitude expressed in the episode, even if occasionally it merely amounts to an awareness of an episode of thinking.

#### **4. Second-order reflection**

Having argued that a modest construal of ET is possible, my next question is whether we have any reason to favour it over an ambitious reading. In the remainder of this paper I want to sketch an argument for an affirmative answer. The argument turns on the interpretation of Evans's 'second-order reflection'. It holds that on closer examination, such reflection gives us a reason to endorse O'Shaughnessy's denial of a 'cognitive path', open to reflection, via which self-knowledge is reached.

As Byrne observes, Evans's remarks 'contain no actual argument'. They just 'strike many as one of those things that are obvious once pointed out.' (2011, 204) It is a good question, though, what makes these remarks so compelling. Note that Evans does not, or not just, provide a description of what it was like for him to reflect on his view of the prospects for world peace. He makes a completely general modal claim. (As in Descartes' Meditations, the first person is used to represent any reflective thinker.) Here is the claim:

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?' (1982, 225, my emphasis)

The 'must' here is naturally interpreted as a 'must' of practical necessity. The point is not, for example, that I 'must be attending' to certain phenomena, in the sense that it is very likely that I do so. Rather the claim is that my intention to state my view gives me a decisive reason to answer the corresponding first-order question. What makes that claim so compelling?

Philosophers who pursue an ambitious construal of ET tend to assume that the answer has to do with the epistemology of self-knowledge. My objective is to let you know what I believe. To do so, evidently, I myself need to acquire *knowledge* of what I believe. That's why I must answer the first-order question. For in doing so I secure the most reliable epistemic basis available to me. On this analysis, that ET strikes us as one of those things that are obvious once pointed out reflects our commitment to

(2). We feel that an answer to an empirical question, even a question about our own current beliefs, stands in need of an epistemic basis; and we take it that the first-order judgement provides just such a basis.

Moran is uneasy about this analysis. He frequently cautions against a ‘too purely epistemic’ conception of the first-person perspective on beliefs. (e.g. 2001, 38)

Importantly, he also provides materials for an alternative account — an account in terms of a distinctive *commitment* incurred by self-ascriptions of belief, rather than in terms of their *epistemic basis*. In his discussion of Moore’s Paradox, Moran writes: ‘in saying “I believe it’s raining out” I commit myself to the state of the weather being a certain way.’ (2001, 74) We can put the idea by saying that it’s a requirement of rationality to *endorse* what one takes to be one’s present beliefs. The ‘Moorean’ judgement ‘I believe that p but not p’ violates that requirement. This may be at least part of the reason the judgement seems intuitively absurd, despite its coherence. Now if there is such a requirement of reflective endorsement (RE), this would provide an alternative explanation of the intuitive force of Evans’s second-order reflection.

Perhaps we feel that it’s on pain of *irrationality* that I ‘must’ address the first-order question about the war. If a self-ascription of belief commits one to the corresponding first-order judgement, the only responsible way to self-ascribe beliefs is by reflecting whether one is in a position rationally to *make* the relevant first-order judgement. To do that just is to address the first-order question. On this analysis, it is not because of any concern with the epistemology of self-knowledge that we are disposed to find ET compelling. Indeed we may not ordinarily devote much attention to our immediate self-knowledge, and it may not occur to us that such knowledge might stand in need

of an epistemic basis. The intuitive plausibility of ET may simply reflect our grasp of the norms of rationality.

At this point, it is natural to wonder whether the two accounts should be seen as mutually incompatible. There may be more than one reason to accept Evans's practical 'must'. Moran's repeated insistence that we should not take 'too *purely* epistemic' a view of ET, I think, reflects precisely such a compatibilist stance. But I want to suggest that there are reasons to be sceptical about compatibilism.

The challenge I see for compatibilism arises from the fact that both RE and the ambitious reading of ET have implications for the basis on which one should form beliefs about one's own beliefs. To comply with RE one needs to look to a basis that properly settles the first-order question. That someone just told me it was raining, for example, can be a good basis for thinking it is raining. If it's a rational requirement to endorse what one takes to be one's current beliefs, I can satisfy that requirement by self-ascribing the belief 'It's raining' on that basis. At the same time, if there is an epistemic basis in terms of which knowledge of our own beliefs can be explained and vindicated, then I should self-ascribe beliefs on *that* basis. Quite generally, if X is the best epistemic basis one has for establishing whether q, surely one should answer the question whether q on the basis of X. The problem is that the fact that someone just told me it was raining is not an intelligible basis for my knowledge that *I believe* it's raining. So I should self-ascribe the belief on a different basis, one that is germane to the psychological question of what I believe. Yet on the face of it, in doing so I am bound to flout RE. I would seem, in other words, to be subject to conflicting requirements. In the next two sections I develop this objection in more detail. I first

elaborate the alternative, ‘Moorean’ account of Evans’s practical ‘must’. In section 6 I present the case for incompatibilism.

### **5. Reflective endorsement**

It seems a natural thought that if someone asks you about your views e.g. on international politics or the weather, you think of yourself as an *informant*. You assume that what, ultimately, the questioner wants to know is whether it’s raining or whether there will be war. So the first question you face is whether you are a useful informant on these matters — whether *you* know what the questioner wants to know. Now in the third world war example, it would be a bit pretentious to make a claim to knowledge. And it might be said that there are areas in which knowledge is simply not a sensible goal. (Aesthetic judgements might be an example.). But even in these cases one can think of oneself as an informant at least in the sense that one can offer a reasoned view, one that the questioner too will have reason to accept or at least to take into consideration. It will be useful to introduce some terminology. Let’s call the question ‘Do I know whether p or do I at least have a (justifiable) view on whether p’ a ‘reflective first-order question’. It is a first-order question in the sense that an affirmative answer to it will, for me, simultaneously settle the question whether p; but it is reflective in that it concerns *my* epistemic or justificatory position vis-à-vis that question. As I said, it seems plausible that we do in fact tend to understand the question of whether we hold a certain belief as a reflective first-order question. But is this mandatory? What would be wrong with reflecting on one’s beliefs as a ‘merely psychological’ matter, removed from any concern with knowledge or reasons?

*That* there would be something wrong with this is suggested by Moore's paradox. In Moran's phrase, a self-ascription of belief 'commits' one to the relevant first-order judgement. And a judgement, in turn, commits one to a claim to knowledge or at least justified belief.<sup>18</sup> But *why* should reflective endorsement in this sense be a rational requirement?

We can approach this by testing the strength of the Moorean intuition. It might be said that it can be a commendable manifestation of *rationality* to be able to adopt a wholly disengaged stance towards one's present beliefs. Suppose you find yourself experiencing recurrent bouts of resentment, and on that basis conclude 'I believe I have been treated unfairly.' When you reflect on the matter, though, you realize that you have been treated exactly the same way as everyone else. So you are not disposed, on reflection, to judge that you have been treated unfairly, and you don't think that you have a good reason for this belief, let alone know it to be true. It's agreed on all hands that your belief is irrational. But it might be said that your detached judgement that *you have the belief* is not. Admittedly the example is underdescribed, and there is of course a significant complication in that your reasoning on the matter itself constitutes relevant evidence: even a well-informed external interpreter would, in the light of it, presumably stop short of an outright attribution of the belief to you. But let's ignore the complication for a moment. Let's suppose that in cases such as this it's the non-verbal evidence — such as affect — that

---

<sup>18</sup> For current purposes it does not matter whether the disjunctive formulation should ultimately be discarded in favour of by a simple account. For the view that knowledge is the constitutive norm of assertion, see Williamson (2000), (2009). For the suggestion that the norm is provided by justified belief, see Kvanvig (2009).

should be treated as decisive. On that supposition, what would be wrong with your self-ascription?

On Moran's account, the answer has to do with the sense in which one may be said to be responsible for one's beliefs. 'The immediacy of avowal', he writes, 'is not only a possibility for the agent but also something of a requirement, in the sense that it is an expression of the person assuming responsibility for this thought and action.' (2001, 131) Your self-ascription, on this view, would be akin to someone's prediction that despite her resolution to stay away from the gambling table, she will most likely succumb to temptation and go back. Now in the latter case, it might be suggested, what makes the prediction irrational is that the agent's intention to quit gambling commits her to the belief that she has the power to do so, yet her prediction suggests she is not convinced on that score. Moran's view seems to be that the belief case is exactly analogous. Your self-ascription of the belief that you have been treated unfairly reflects the view that your belief on this matter is not responsive to reason. Yet insofar as you *deliberate* about the issue at all you are inevitably committed to taking it to be within your power to determine what you believe by reflection on the relevant reasons.<sup>19</sup>

One might question the proposed analysis of the gambling case. But even supposing that it is on the right lines, it is not clear that it offers a helpful model for the belief case. There are, after all, significant disanalogies. Part of what underpins our judgement that the gambler's prediction is irrational is arguably that she *formed* the

---

<sup>19</sup> 'Indeed, there is no point in calling it "deliberation" any more, if [the deliberator] takes it to be an open question whether this activity will determine what he actually does or believe. To engage in deliberation in the first place is to hand over the question of one's belief or intentional action to the authority of reason.' (2001, 127)



intention to quit gambling. If she does not believe she will be able to quit, then, put bluntly, she shouldn't have formed that intention. In the belief case, the putative practical commitment to being an effective deliberator does not come up for review, as it were: it's supposed to be something we are simply saddled with, for so long as we *are* reflective thinkers. It seems to me that this affects the role that this commitment can play in our assessment of the rationality of your evidence-based self-ascription. Certainly it seems less obvious than in the gambling case that your practical commitment means the detached self-ascription ('I believe I have been treated unfairly') must be irrational. Given that you haven't, and couldn't possibly have, *undertaken* the commitment, there is some uncertainty as to whether you should be held to it.<sup>20</sup>

This is not intended as a decisive objection to Moran's account. But I want to suggest that there is an alternative, and more straightforward, way to understand the sense in which your self-ascription would be irrational. Attributions of belief have a certain 'normative import'.<sup>21</sup> If A believes that p, this has implications for what it would be rational for A to do and believe, given his other attitudes. Now in the case of attributing beliefs to others (and in the case of past-tense self-ascriptions) these normative implications are fairly minimal. Suppose A believes that p and the truth of p would give him a reason to  $\phi$ . Then there is a sense in which it would be rational for

---

<sup>20</sup> In the light of this, it is perhaps unsurprising that Moran's own position here is not easy to pin down. Shoemaker interprets him as *denying* that a Moorean self-ascription must be irrational. Someone undergoing psychoanalysis might, according to Shoemaker's Moran, actually come to *know* that he has a certain belief merely on the basis of evidence, despite regarding the belief as false and irrational. Shoemaker objects to this that the 'Moore-paradoxical sentence "I believe that my brother betrayed me, but he didn't" remains paradoxical when it is uttered by Moran's analysand.' (2003, 393) As Shoemaker acknowledges, though, some of what Moran says is hard to square with this reading.

<sup>21</sup> I borrow the phrase from Millar (2004).

A to  $\phi$ .<sup>22</sup> We might even advise A, in a ‘subjective’ key as it were, that in the light of what, from his point of view, seem to be the relevant considerations, *he should*  $\phi$ . But that’s of course consistent with advising him, in an ‘objective’ key, that as *we* see things, it is not true that he should  $\phi$  — he has no reason to  $\phi$ , given that, as we know or believe, it’s not true that  $p$ .<sup>23</sup> There is evidently no conflict between the two kinds of advice: ‘objective’ advice reflects the (current) beliefs of the adviser, ‘subjective’ advice is relativized to the advisee’s beliefs. But now consider a first-person self-ascription of belief, such as your judgement ‘I believe I have been treated unfairly’. This commits you to advising yourself, ‘subjectively’, that you should do certain things (those that the truth of the belief would give you reason to do). At the same time, given your judgement that the belief is false and irrational, you should advise yourself, ‘objectively’, that you have no reason to do them. The problem is that in the first-person case, the two kinds of advice really are in conflict: while your ‘objective’ advice reflects how *you* see things, your ‘subjective’ advice is relativized to that same perspective, viz. yours. You’d be committed to both affirming and denying that you should act in certain ways. If this suggestion is on the right lines, we can see that reflective endorsement is a rational requirement by reflection on the normative commitments inherent in attributions of belief, quite independently of any demands that may be held to be implicit in the perspective of the deliberator.

## 6. Incompatibilism

---

<sup>22</sup> That there is some such sense is arguably consistent not just with the falsity but even with the irrationality of A’s belief that  $p$ : see Bratman (1987) for illuminating discussion of the different points of view from which we assess the rationality of actions.

<sup>23</sup> For the distinction between ‘subjective’ and ‘objective’ advice, see Kolodny (2005).

Suppose that self-knowledge, contra O'Shaughnessy, has an epistemic basis in terms of which the subject may find her possession of it intelligible. It would follow that the subject should form beliefs as to what she believes on that basis. That would be the rational way to go about reflecting on one's beliefs. But in fact such a belief-forming practice would be notably *irrational*. For it would violate the requirement of reflective endorsement. It would disable the subject from thinking of the question of what she believes as the question of whether she knows or has a justified view.

We can distinguish two strategies compatibilists might deploy to resist this conclusion. One would be to locate the epistemic basis for self-knowledge in the very considerations on the basis of which we make a claim to knowledge or reasoned belief. The idea would be, for example, that I answer the question of whether I believe that it's raining on the basis of the facts that constitute my reason to believe it *is* raining. This is the natural way to read Moran's suggestion that ET amounts to 'a claim about *how* a set of questions is to be answered, what sorts of reasons are to be taken as relevant'. Specifically, his claim is (recall) that 'a first-person present tense question about one's belief is answered by reference to (or consideration of) the same reasons that would justify an answer of the corresponding question about the world.' (2001, 62) Suppose my reason for believing it is raining is the fact that you just said it was. Doubtless that fact can be a conclusive basis for thinking it is raining, but it is not much of a reason to think that I believe it is raining. What's more, I can know what I believe quite independently of whether you really said it was raining. Perhaps I misheard you. In that case, on the face of it, I have no reason that 'would justify an answer to the corresponding question about the world'. But I could still, in that situation, know that I believed it was raining.

It's natural, at this point, to shift to an alternative reading of Moran's proposal. Perhaps when he refers to 'reasons' he doesn't really mean justifying reasons, i.e. facts, but the sorts of 'reasons' appealed to in 'rationalizing explanations', i.e. beliefs. Then the basis of my self-ascription would be my *belief* that you just said it was raining, and perhaps my taking your utterance to be a good reason to think it's raining.

Note, though, that if this is how the suggestion is to be understood, it is misleading to say that the second-order question is answered by reference to the '*same reasons*' that justify the answer to the first-order question. I do not answer the first-order question by reference to my believing you told me it was raining, or to my taking this to be a good reason, but simply by reference to the fact that you told me and that this is a good reason.<sup>24</sup> So our re-interpretation of Moran's proposal really marks a shift to an alternative compatibilist strategy. The idea now is that I answer the first-order question on *one* sort of basis (e.g. that you said it was raining), and that this, in turn, provides me with *another* kind of basis (my reflection on the reasons) on which to self-ascribe the belief. The two bases are distinct but, you might insist, essentially in harmony.

The problem, I suggest, is that the second basis comes too late to do any useful work. A responsible answer to the first-order question will need to be based on considerations that convince me that I can serve as an informant on this matter, i.e.

---

<sup>24</sup> As Barry Stroud reminds us, 'My *believing* that it has always rained [in Berkeley] does not amount to much of a reason for expecting rain in Berkeley in January. In trying to determine what to believe, or what there is most reason to believe, my focus is on the world, or on what is so, not on my attitudes toward what is so.' (1999, 34)

that I know the answer or at least have a justifiable view. I cannot, in other words, separate the first-order question from what I called a *reflective* first-order question. I cannot settle the one without settling the other. But once I have reached the view that I know, or at least have a justifiable belief, that it's raining, it is no longer an open question for me what I believe on this subject. There would seem to be no need to look to any further epistemic basis to address the latter question. Indeed there can be no room for any such basis. I'm already committed to the relevant self-ascription of belief, in virtue of my answer to the (reflective) first-order question.

In summary, neither of the two compatibilist strategies — invoking a single epistemic basis vs two distinct but intelligibly connected bases — holds much promise. I suspect the intuitive appeal of Moran's picture derives from an unnoticed oscillation between the two readings. It can seem intuitive that we self-ascribe beliefs on the basis of our first-order reasons. Those reasons, however, cannot provide an intelligible basis for self-knowledge. So we re-interpret our intuition, appealing to a more 'internal' basis: our *reflection* on our first-order reasons. But it's much less plausible that we actually use that alternative basis, and indeed that we could rationally do so.<sup>25</sup>

## 7. Conclusion

---

<sup>25</sup> I think we can observe a similar oscillation in Byrne's discussion. The suggestion that I 'infer' that I believe that p from the premise that p has intuitive appeal insofar as it captures the sense I have that whether I believe that p turns on whether p. But in the light of Byrne's explanation of what makes this type of 'inference' knowledge-conducive it turns out that the truth of my premise is totally irrelevant to the truth-conduciveness of the 'inference'. That explanation suggests the real basis of my self-knowledge is not the fact that p but my *believing* that p. For what explains the truth-conduciveness of the 'inference' is that 'inference from a premise entails belief in the premise'.

My concern in this paper has been the question whether knowledge of one's current beliefs is intelligible to the subject in terms of some epistemic basis or is acquired 'silently'. I have qualified O'Shaughnessy's blanket rejection of internal intelligibility by arguing that there is room for a weak interpretation of ET. On this interpretation, we know our current beliefs in, rather than on the basis of, making judgments about the world. This would enable the subject to be aware, in a sense, of how self-knowledge is obtained (viz. by answering questions about the world). But I have sketched an argument in support of the thesis that I suspect was O'Shaughnessy's main point, his denial that there is an epistemic basis, open to reflection, in terms of which one's possession of self-knowledge can be seen to be intelligible.

How damaging would that conclusion be to the prospect of an ambitious construal of ET? It might be said that my argument only affects 'internalist' versions of the ambitious approach, on which the basis for self-knowledge is one the subject herself may invoke in defending her second-order beliefs and explaining her possession of self-knowledge. Thus a concessive response to the argument might be that this kind of internalism would indeed be inconsistent with the characteristic commitments of self-ascriptions of belief, but that we can still substantially advance our philosophical understanding of self-knowledge through an analysis of the basis on which second-order beliefs, insofar as they are acquired by the use of Evans's 'procedure', are formed. I conclude by raising a question about the stability of this line of response.

Note that the account would be 'externalist' not just in the sense that the subject lacks a proper understanding of what makes the basis on which her second-order beliefs are acquired 'knowledge-conducive', but in the more radical sense that the subject is not

even aware of the existence of any such basis. This is not necessarily a problem. We are familiar with externalist theories of perceptual knowledge that invoke the reliability of non-conscious information-processing mechanisms in accounting for the ‘status as knowledge’ of beliefs generated by them. The question I want to raise is how we are supposed to *know* that second-order beliefs are actually acquired on the putative basis. This is a substantive claim about the causal explanation of certain beliefs. The analogous claim about non-conscious perceptual mechanisms is of course part of an empirical theory, supported by evidence. But what grounds do we have for thinking that in using Evans’s ‘procedure’ we (typically) acquire a second-order belief that is be causally explained in terms of a certain kind of epistemic basis (say, our first-order judgment)?

I think there is a real problem here. If we adopt a radically externalist approach, we should not expect to be able to discover the basis of second-order beliefs simply through reflection on what we intentionally do when we reflect on our own beliefs, any more than we should expect to discover the non-conscious mechanisms underpinning vision by reflectively looking at the world. We should not expect to find that we ordinarily use a certain kind of basis as a means for the acquisition of self-knowledge. If ordinary reflective thinkers intentionally used some such basis, they could hardly be completely unaware of its existence.<sup>26</sup> Here we have come full circle.

---

<sup>26</sup> My suggestion, in other words, is that it is no accident that extant ambitious construals of ET encourage the thought that the epistemic basis for self-knowledge is open to reflection. To illustrate, if, as Byrne thinks, I *infer* that I believe that p from the single premise that p, it is natural to expect that I can be aware that this is what I am doing. Moran’s account, recall, is explicitly framed as a solution to the puzzle of how it can ‘*make sense for me*’ to follow Evans’s procedure, in the face of the disparate subject matters of the two questions involved. (See above, n. 8) This is a question about an intentional activity of ordinary reflective subjects. If the puzzle has a solution and the activity does after all make sense, presumably this means it makes

What has made ET look so attractive to several generations of philosophers is that it appears to offer a compelling articulation of what we do when we self-ascribe beliefs. The externalist approach would be debarred from interpreting ET in this way. It would therefore raise the question of how it is possible, from Evans's armchair, to discover an epistemic basis that is not open to reflection.

### **Acknowledgements**

Previous versions of this paper were presented at workshops or seminars in Madrid, Salzburg, London and Buenos Aires. I am grateful to participants for their comments. Special thanks to two anonymous referees for helpful criticism, and to Bill Brewer, Stephen Butterfill, Naomi Eilan, Christoph Hoerl, Diego Lawler, Hemdat Lerman, Guy Longworth, and Matt Soteriou for numerous discussions about self-knowledge.

### **References**

- Austin, J.L. 1961. Other Minds. In his *Philosophical Papers* (ed. by J. Urmson and G. Warnock), 76-116. Oxford: Clarendon Press.
- Boghossian, P. 2008. Content and Self-Knowledge. In his *Content and Justification*, 139-158. Oxford: Oxford University Press.

---

sense *to* those engaging in it. Moran's proposed solution, in terms of the content of the conception we have of ourselves as deliberators, suggests he accepts this constraint.



- Boyle, M. 2009. Two Kinds of Self-Knowledge. *Philosophy and Phenomenological Research* 78: 133-164.
- Bratman, M. 1987. *Intentions, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.
- Byrne, A. 2011. Transparency, Belief, Intention. *Proceedings of the Aristotelian Society, Supplementary Volume LXXXV*: 201-221.
- Dennett, D. 1991. *Consciousness Explained*. London: Allen Lane.
- Evans, G. 1982. *The Varieties of Reference*. Oxford: Clarendon Press.
- Gallois, A. 1996. *The World Without, the Mind Within*. Cambridge: Cambridge University Press.
- Hampshire, S. 1979. Some Difficulties in Knowing. In T. Honderich & M. Burneyat (eds.), *Philosophy As It Is*, 281-308. London: Allen Lane.
- Heal, J. 2003. On First-Person Authority. In her *Mind, Reason, and Imagination*, 273-288. Cambridge: Cambridge University Press.
- Hobson, P. 2002. *The Cradle of Thought*. London: Macmillan.
- Kolodny, N. 2005. Why Be Rational? *Mind* 114: 509-563.
- Kvanvig, J. 2009. Assertion, Knowledge, and Lotteries. In P. Greenough and D. Pritchard (eds.), *Williamson on Knowledge*, 140-160. Oxford: Oxford University Press.
- Martin, M. 1998. An Eye Directed Outward. In C. Wright, B. Smith & C. MacDonald (eds.), *Knowing Our Own Minds*, 99-121. Oxford: Oxford University Press.
- McDowell, J. 1982. Criteria, Defeasibility, and Knowledge. *Proceedings of the British Academy* 68: 455-79.
- Millar, A. 2004. *Understanding People*. Oxford: Clarendon Press.
- Moran, R. 2001. *Authority and Estrangement*. Princeton: Princeton University Press.

- Moran, R. 2003. Responses to O'Brien and Shoemaker. *European Journal of Philosophy* 11: 402-19.
- Moran R. 2005. Problems of Sincerity. *Proceedings of the Aristotelian Society* 105: 235-45.
- O'Shaughnessy, B. 2000. *Consciousness and the World*. Oxford: Oxford University Press.
- Peacocke, C. 1996. Conscious Attitudes, Attention, and Self-Knowledge. In C. Wright, B. Smith & C. MacDonald (eds.), *Knowing Our Own Minds*, 63-98. Oxford: Oxford University Press.
- Reed, B. 2010. Self-Knowledge and Rationality. *Philosophy and Phenomenological Research* 80: 164-81.
- Shaftesbury (Anthony Ashley Cooper, Third Earl of Shaftesbury). [1711] 1999. *Characteristics of Men, Manners, Opinions, Times*. Cambridge: Cambridge University Press.
- Shoemaker, S. 2003. Moran on Self-Knowledge. *European Journal of Philosophy* 11: 391-401.
- Stroud, B. 1999. Practical Reasoning'. In E. Ullmann-Margalit (ed.), *Reasoning Practically*, 27-38. New York: OUP. 27-38.
- Williams. B. 2002. *Truth and Truthfulness*. Princeton: Princeton University Press.
- Williamson, Timothy 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
- Williamson, Timothy 2009. Reply to Jonathan Kvanvig. In P. Greenough and D. Pritchard (eds.), *Williamson on Knowledge*, 341-349. Oxford: Oxford University Press.

