

## EVANS ON SELF-KNOWLEDGE

Quassim Cassam

[Draft of a paper to be published (in German) in C. Misselhorn (ed.) *Sprache, Wahrnehmung und Objektivität: Neue Perspektiven auf die Philosophie von Gareth Evans*]

1

In a famous passage Gareth Evans discusses our ways of knowing what we think or believe. This is the passage, from chapter 7 of *The Varieties of Reference*:

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?’. I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*..... If a judging subject applies the procedure, then necessarily he will gain knowledge of one of his own mental states: even the most determined sceptic cannot find here a gap in which to insert his knife. We can encapsulate this procedure for answering questions about what one believes in the following simple rule: whenever you are in a position to assert that *p*, you are *ipso facto* in a position to assert ‘I believe that *p*’ (1982: 225-6).

Evans adds that mastery of this procedure cannot constitute a full understanding of the content of the judgement ‘I believe that *p*’. The subject must also be able to conceive of subjects other than himself as believing that *p*.

Many questions have been raised about Evans’ proposal in the extensive secondary literature it has generated. As he admits, his discussion is ‘extremely incomplete’ (1982: 224), and some of the questions that have been raised about his proposal are a reflection of that fact. Nevertheless, it has seemed to most commentators that Evans is saying something true and important about knowledge of our own beliefs. The challenge is to figure out what this important truth amounts to. This involves determining, among other things, what Evans’

theory is a theory *of*: is it an account of how we *do*, how we *can*, or how we *must* come to know our own beliefs?

I'll be arguing for two main claims here:

(A) While Evans identifies a *possible* route to knowledge of our own beliefs, it's not clear either that we *must* acquire knowledge of our own beliefs by executing the procedure he describes or that we *do* generally come to know our own beliefs in this way.

(B) The self-knowledge that Evans describes is *indirect* self-knowledge. He describes a procedure for answering questions about what one believes, and it's hard to make sense of this procedure other than on the assumption that the knowledge it produces is neither epistemologically nor psychologically immediate.

What (A) suggests is that Evans' account is much more limited in scope than is sometimes supposed. One limitation is that his account is an account of how one is able to know whether one believes *p* in cases in which it is explicitly in question whether one believes *p*.<sup>1</sup> Suppose we call these 'in question' (IQ) cases. It should be noted that it is also possible to come to know or realize that one has a certain belief even when the question whether one has that belief hasn't been explicitly raised. Even if it isn't in question whether you believe *p* you may suddenly realize, perhaps as a result of hearing or reading a compelling formulation of *p*, that you do in fact believe it. It isn't always clear in such cases whether your belief is newly formed, one you already had, or something in between. Evans doesn't account for such cases. As we will see, there are also questions about his account of self-knowledge in IQ cases.

The significance of (B) is that philosophers of self-knowledge often start their discussions by commenting on what makes self-knowledge epistemologically distinctive when compared with knowledge of others. A standard intuition is that self-knowledge is normally direct, both psychologically and epistemically, whereas knowledge of others is

normally indirect. As far as self-knowledge is concerned, the psychological claim is that you normally know your own beliefs without any conscious reasoning or inference. You normally know your own beliefs *immediately*: you don't have to *do* anything in order to know what you believe, you just know. The epistemological claim is that your knowledge of what you yourself believe is normally non-inferential and not based on evidence, behavioural or otherwise. If you know that you believe that P, then you must be justified in believing that you believe that P. In these terms, the idea is that your justification for believing that you believe that P doesn't come from your justification for believing any other proposition.<sup>2</sup>

Surprisingly, Evans' account of self-knowledge has been endorsed by philosophers who regard the directness of ordinary self-knowledge as a datum.<sup>3</sup> Such philosophers have tried to represent Evans' procedure as delivering direct, non-inferential self-knowledge but (B) implies that this is a mistake. You can't have it both ways: if self-knowledge is normally direct then executing Evans' procedure can't be the normal way of getting it. I don't claim that this is how Evans saw things. His discussion isn't organized around a distinction between direct and indirect knowledge, and the extent to which he saw his account as inferential is hard to know. (B) is a claim about what is the case rather than about Evans' conception of his own theory. Since, unlike many philosophers of self-knowledge, I don't regard the directness of self-knowledge as a datum, I don't mind the fact that Evans' story is an inferential story.<sup>4</sup> My reservations have other sources and centre on (A) rather than (B).

## 2

Let's concentrate to begin with on how we come to know our own beliefs in IQ cases. It's revealing how often Evans characterizes his account as an account of a 'procedure'. The dictionary definition of a procedure is a way of proceeding, a mode of performing a task or 'a series of actions conducted in a certain order or manner'. If you answer the question whether you believe there will be a third world war by putting into operation a procedure in the latter

sense then resulting self-knowledge (if any) is psychologically indirect. It's certainly not the case on Evans' account that you don't have to do anything in order to answer the question, or that you know straight off what you believe. In addition, the procedure you supposedly have to put into operation in order to answer the question whether you believe  $p$  is the procedure for answering an apparently *different* question, the question whether  $p$ .<sup>5</sup> How can the self-knowledge acquired by putting into operation a procedure for answering a different question be anything other than indirect? There must be some kind of transition from answering the question whether  $p$  to answering the question whether you believe that  $p$ , but knowledge that is genuinely direct would involve no such mental transition.

One reason this seems problematic is that it seems a poor fit with cases in which you have already made up your mind (perhaps by reasoning) whether  $p$ , and in which the question whether you believe that  $p$  is naturally heard as the question whether you *already* believe that  $p$ .<sup>6</sup> In such cases you don't have to put into practice whatever procedure you have for answering the question whether  $p$ .<sup>7</sup> Since you've already made up your mind, all you have to do is retrieve a pre-existing belief. This will typically simply be a matter of recalling your view, perhaps instantaneously, rather than performing a series of actions conducted in a certain order or manner.

This raises a question about the status of Evans' theory. As we have seen, his view is that if someone asks me whether I believe that  $p$  I *must* attend, in answering him, to the same outward phenomena as I would attend to if I were answering the question whether  $p$ . But why must I attend to the outward phenomena if my mind is already made up? One way of putting this would be to distinguish between two IQ cases, cases of pre-existing belief (I'll call them PEB cases) and 'making up your mind' cases (I'll call them MUM cases). In a MUM case you don't already believe that  $p$ . Instead, the question 'do you believe that  $p$ ?' is heard as an invitation to make up your mind whether  $p$ , that is, to form a belief with respect to  $p$ . In PEB

cases, you do not have to make up your mind because your mind is already made up. In these terms, one might think that Evans' account is, at best, an account of how the question 'do you believe that  $p$ ?' is or must be answered in MUM cases. It is not an account of how we must or do come to know our own beliefs in PEB cases, or in non-IQ cases.

This way of reading Evans raises two questions:

1. Is it true that it doesn't work for PEB cases?
2. Does it even work for MUM cases?

With regard to the first issue, Richard Moran observes that while I can raise the question of what *another* person believes about  $p$  'without that person considering the question of the truth of  $p$  itself' (2012: 223), this is not possible when there is only one person on the scene: 'I cannot pose to myself the question whether I believe that  $p$  without raising the question of the truth of  $p$ , for there is only one mind under consideration here, inquiring about itself' (ibid.). A belief of mine only counts as a genuine *belief* if I take it to be true, and this applies with equal force to pre-existing, stored beliefs. A stored belief cannot therefore be 'insulated from the engagement of my rational capacities for determining what is true or false' (2012: 222).

Although there is something right about Moran's observation, its epistemological significance is unclear. One's beliefs can't be insulated from one's rational capacities in the sense that a person who believes a certain proposition can be asked why they believe it. This is a request for their *reasons*, and is a reflection of the fact that 'the asking and giving of reasons belongs to the nature of belief itself' (Moran 2012: 216). However, it doesn't follow that one comes to *know* one's stored beliefs by reflecting on one's reasons, or that we come to know our beliefs in PEB cases in the way that we come to know our beliefs in MUM cases. We lack the time and mental resources to rethink the question whether  $p$  every time we are asked whether we believe that  $p$ . In cases of pre-existing belief, there certainly needs to be a

*preparedness* to rethink one's belief if relevant new evidence comes to light but being prepared to raise the question of the truth of *p* is not the same as actually raising this question. Sometimes there is no such question (at least in one's own mind) and one is simply able to report one's pre-existing belief without executing Evans' procedure.

It's one thing to say that Evans fails to account for PEB cases but does he at least give a plausible account of self-knowledge in MUM cases? The difficulty is this: suppose that you haven't already made up your mind whether *p* and the question arises whether you believe that *p*. So you follow Evans' advice and put into operation your procedure for answering the question whether *p*. As a result you judge that *p*. How do you get from there to knowing that you believe that *p*? Your procedure for answering the question whether *p* might lead you to judge that *p*, but what is the connection between *judging* that *p*, *believing* that *p* and *knowing* that you believe that *p*? Evans doesn't say, and this makes it difficult to know what his account of self-knowledge amounts to.

At this point there are two ways of proceeding. One approach insists that there is a gap between judging that *p* and knowing that you believe that *p* and that this gap can only be bridged by some kind of inference. A different approach denies the existence of any such gap, and argues that when you put Evans' procedure into operation in a MUM case you come to know *without inference* that you believe that *p*. There are many different versions of each of these approaches and I want to argue for a version of the first approach. The implication of my proposal is that in MUM cases it is possible to acquire self-knowledge roughly in the way that Evans describes. Whether, even in such cases, we do or must proceed in this way is a further question. What should not be in question, in my view, is that when you put Evans' procedure into operation in a MUM case the resulting self-knowledge is inferential.

As a way of working up to a defensible account of self-knowledge in MUM cases let us first briefly consider Alex Byrne's version of inferentialism. I think that the difficulties

with Byrne's proposal point the way to a better inferentialist approach. Responding to Evans, Byrne writes:

Suppose that I examine the evidence and conclude that there will be a third world war. Now what? Evans does not explicitly address this question, but the natural answer is that the next step involves an *inference from world to mind*: I infer that I believe that there will be a third world way from the single premiss that there will be one (2011: 203).

Byrne calls the inference from world to mind a 'transparency inference'. The question 'do I believe that  $p$ ?' is 'transparent' to the question whether  $p$  in the sense that the former question is answered or answerable by answering the latter. Byrne represents transparency inferences as closing a gap in Evans' account but this proposal faces a serious objection: transparency inferences in Byrne's sense are patently invalid.<sup>8</sup> The mere fact that  $p$  is true doesn't entail that I believe it so how can it be legitimate to infer 'I believe that  $p$ ' from the fact that  $p$ ?

Byrne argues that this difficulty isn't insuperable and that transparency inferences are not without epistemic merit. If one infers that one believes that  $p$  from the single premiss  $p$  then 'one's second-order belief is *true*, because inference from a premiss entails belief in that premiss' (2011: 206). In addition, transparency reasoning 'typically yields beliefs that are *safe* in the sense that they could not easily have been false' (2011: 206-7). Taken together, these considerations suggest that transparency reasoning is knowledge-conducive. However, Byrne is wrong to assume that inference from a premiss entails belief in that premiss. Logic teachers run thousands of sample inferences from premisses that neither they nor anyone in their right mind actually believes. In *reductio* arguments, one supposes that  $p$ , infers  $q$  from  $p$ , and then infers the falsity of  $p$  from the falsity or absurdity of  $q$ . There is obviously no question here of inference from a premiss entailing belief in that premiss. In Byrne's scenario, therefore, the inference from ' $p$ ' to 'I believe that  $p$ ' will only be legitimate if you

know that the premiss gives expression *belief* of yours rather than a supposition or conjecture. But now it seems that knowledge of what you believe is being presupposed rather than explained.<sup>9</sup>

These considerations don't show that there is anything wrong with inferentialism *per se*, only that there is something wrong with Byrne's version of inferentialism. Since he sees transparency inferences as unmediated, as moving directly from '*p*' to 'I believe that *p*', an alternative would be to think of such inferences as *mediated*. As an illustration, suppose once again that in response to 'Do you think there is going to be a third world war?' you examine the evidence and conclude that there will be a third world war. To 'conclude' that there will be a third world war is to *judge* that there will be one. What, then, is the relationship between judging and believing? Judging, it is often said, is the formation of belief, but it is possible to judge that *p* without believing that *p*. Consider this example:

Someone may judge that undergraduate degrees from countries other than their own are of an equal standard to her own, and excellent reasons may be operative in her assertions to that effect. All the same, it may be quite clear, in decisions she makes on hiring, or in making recommendations, that she does not really have this belief at all (Peacocke 1998: 90).

How is it possible for a person to judge that *p* and yet fail to believe that *p*? Suppose we think of judgement as a cognitive mental *act*, the act of occurrently putting a proposition forward in one's mind as true, and belief as a mental *state* or cognitive attitude.<sup>10</sup> In a given case, the judgement that *p* might fail to lead to the belief that *p* because belief-formation is also influenced by non-rational factors such as self-deception, prejudice and phobias.<sup>11</sup> I might judge for good reasons that undergraduate degrees from countries other than my own are of an equal standard to my own and yet find myself unable to take this to heart as a result of a prejudice which I just can't shake off. I mentally *affirm* that undergraduate degrees from other



countries are of an equal standard and yet my attitude towards this proposition is not the attitude of belief, as evidenced by my hiring decisions and letters of recommendation.<sup>12</sup>

Even if judging that  $p$  doesn't always lead to the belief that  $p$ , it is nevertheless plausible that judging that  $p$  normally one leads one to believe that  $p$ . The formation of the belief might be blocked by a prejudice or phobia but this can hardly be the normal case. It's hard to conceive of a subject who judges that  $p$  but hardly ever comes to believe that  $p$  as a result. Other things being equal, one would expect someone who genuinely judges that  $p$ , puts the proposition forward in his mind as true, to come to believe that proposition if she does not already believe it. Other things are not always equal, and that is why judgement does not always lead to belief. However, if someone never or hardly ever believes that  $p$  as a result of her supposed judgement or affirmation that  $p$  then it would be reasonable to wonder whether her affirmation is really a judgement rather than, say, a conjecture.

Bearing all these points in mind we can now return to Evans' example. The question arises whether you believe that there will be a third world war so you examine the evidence and judge that there will be one. Let's assume also that you *know* that this is what you judge. Then you can infer that you believe that there will be a third world war as long as you are entitled to assume that, other things being equal, what you judge is what you believe.<sup>13</sup> In this way you can come to know that you believe that there will be a third world war but your knowledge is inferential in the following sense: it is mediated by your knowledge of what you judge and by an assumption about how your judgements and beliefs are linked. In effect, your mental affirmation that  $p$  is serving as evidence that you believe that  $p$ . The evidence isn't infallible since your affirmations aren't an infallible guide to your underlying mental state of belief.<sup>14</sup> In Peacocke's example there is a mismatch between what you judge and what you believe but such mismatches are sufficiently exceptional for you to be entitled to infer from your judgement that there will be a third world war that this is also what you

believe. This is self-knowledge by inference but not in Byrne's sense. You don't infer 'I believe that  $p$ ' from the sheer fact that  $p$  but from your knowingly judging that  $p$ .

Here are five pertinent questions about this model of self-knowledge, which I'll call the Mediated Inference Model (MIM), together with an indication of how I think these questions should be answered:

- i. Is MIM an accurate representation of Evans' view? *Answer:* it's hard to know on the basis of his very brief discussion what Evans had in mind or how he would have closed the apparent gap in his account. The proposal is not that Evans actually thought of mediated inferences as the basis on which one can know what one believes in MUM cases but that this is what he should have thought.
- ii. You can know *on the basis of* judging that  $p$  that you believe  $p$  without *knowing* that you judge that  $p$ , or *inferring* that you believe  $p$  from the premiss that you have judged that  $p$ .<sup>15</sup> In that case, in what sense is your knowledge that you believe  $p$  genuinely inferential? *Answer:* the fact that you judge that  $p$  can't, on its own, justify the belief that you believe that  $p$  unless you *realize* or *know* (and so are justified in believing) that you judge that  $p$  and that, other things being equal, what you judge is what you believe. This makes your knowledge that you believe  $p$  inferential in an epistemological sense: your justification for believing that you believe that  $p$  comes from your having justification to believe other, supporting propositions.<sup>16</sup> Perhaps you aren't *conscious* of inferring your beliefs from your judgements but this leaves the epistemological issues wide open and, in any case, inferences don't have to be conscious.
- iii. MIM presupposes that you know your own judgements but how is *that* self-knowledge possible? Without an answer to this question the proposed explanation of how one knows one's own beliefs in MUM cases looks seriously incomplete.

*Answer:* yes, an account of how one knows one's own judgements is needed. One possibility is that your knowledge of your own judgements is unmediated, direct knowledge. You know, without inference, your own judgements. This is perfectly compatible with MIM, which only says that knowledge of your own *beliefs* is inferential. A different option is to argue that knowledge of your own judgements is also inferential. Here is how: suppose you consider the evidence in favour of *p* and say yourself '*p*'.<sup>17</sup> Have you judged that *p*? It can happen that as you say '*p*' you experience a feeling a cognitive unease, a sense that it doesn't ring true that *p*. You feel that the question is not settled and that you need to think again. In that case, it has *occurred* to you that *p* but you haven't mentally *affirmed* – that is, judged - that *p*. You have only judged that *p* if you have a sense of cognitive ease or settledness when it occurs to you that *p*, a sense of the question being settled in your own mind and no pressing desire to revisit the question. And it is on the basis of such 'internal promptings' that you know that you genuinely *judge* that *p*.<sup>18</sup> You might not consciously infer from such promptings that you judge that *p* but your belief that you judge that *p* is justified by such promptings, and the knowledge to which your belief gives expression is to this extent epistemically indirect.

- iv. In MIM, what entitles you to assume that other things being equal what you judge is what you believe? *Answer:* your grasp of the relevant concepts. You can't think that you believe *p* unless you have the concept of belief. If you have this concept then you understand, perhaps only implicitly, how what you believe is related to what you judge. Like your knowledge of other peoples' beliefs, your knowledge of your own beliefs is grounded in an implicit grasp of a simple theory of belief, a theory of what it is to believe.

- v. Isn't MIM blatantly at odds with the assumption that self-knowledge is normally direct? *Answer:* yes it is, but this is not necessarily a reason for rejecting MIM. It is rather a reason for questioning the assumption that self-knowledge is normally direct. After all, it is possible to think that *some* self-knowledge is normally direct (knowledge of one's own sensations perhaps) without thinking that knowledge of one's beliefs and other standing attitudes is normally direct. To believe that  $p$  is, among other things, to be disposed to think that  $p$  when the question arises, act as if  $p$ , and rely on  $p$  as a premiss in reasoning.<sup>19</sup> To think that knowledge of one's own beliefs is normally direct is to think that it is possible to know directly, without evidence or inference, that one has the relevant dispositions. It isn't at all clear how this is possible. Far from being a datum, the claim that knowledge of one's own beliefs is normally direct is a substantive philosophical thesis that is called into question by MIM, at least on the assumption that MIM offers a plausible account of how we normally come to know our own beliefs in MUM cases.<sup>20</sup>

3

In the discussion so far I've focused on inferentialist readings of Evans. I've argued that Evans' account of how we know our own beliefs is amenable to an inferentialist reading, that there is no objection in principle to inferentialism, and that inferentialism about self-knowledge and dispositionalism about belief are natural allies. However, before going any further down this path, it would be worth pausing to consider whether it is either necessary or desirable to develop Evans' insight along inferentialist lines. I've talked about a gap between judging or believing that  $p$  and knowing that you believe that  $p$  but it has been argued that there is no such gap, and hence no need to posit an implicit or explicit inference to bridge the gap. On this view, whether or not it is a philosophical *datum* that knowledge of one's own

beliefs is normally non-inferential, it is *true* that knowledge of one's own beliefs is normally non-inferential.

Something along these lines is suggested by Matthew Boyle's 'reflectivist' alternative to inferentialism. Boyle agrees with Evans and others that there is '*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 in large' (2011: 227). A philosopher who wants to do justice to this idea while avoiding inferentialism has some options:

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 know 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 (2011: 227).

The reflectivist approach offers an explanation of self-knowledge that is metaphysical rather than epistemological. It denies that in the normal 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' (2011: 235). Accordingly:

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 case

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  $p$  (2011: 228).

The reflectivist operates with an ‘activist’ rather than a dispositionalist conception of belief.<sup>21</sup> Believing  $p$  is actively holding  $p$  to be true. To put it another way, my believing  $p$  is my knowingly evaluating  $p$  as true, and that is why I can treat the question whether I believe that  $p$  as tantamount to the question whether  $p$ : my answer to whether  $p$  expresses my knowing evaluation of  $p$  as true and ‘unless I am in an alienated condition, my knowingly evaluating  $p$  as true just is my believing it’ (2011: 236).<sup>22</sup>

Is it plausible that believing  $p$  and knowing oneself to believe  $p$  are not two cognitive states? As Boyle recognizes, a major worry about his proposal is that it ‘attributes to subjects an implausible omniscience about their own beliefs (implying that whenever one does believe  $p$ , one knows oneself to believe  $p$ )’ (2011: 229). Suppose you believe that  $p$  but for some reason your belief isn’t consciously accessible to you. Isn’t this a case in which you have a belief you don’t know yourself to have? Not according to Boyle. He argues that ‘when a belief is present but not consciously accessible to you, *so too is knowledge of that belief*’ (2011: 229). In other words, we don’t have a case here in which you believe that  $p$  without knowing that you believe that  $p$ . Rather, you know that you believe that  $p$  but you aren’t *aware* of knowing that you believe that  $p$ .

There remains a sense, on this account, in which subjects *are* omniscient about their own beliefs, even if such omniscience is made more palatable by the concession that their knowledge might not be consciously accessible. Regardless of whether one’s self-knowledge is consciously accessible it remains the case for Boyle that whenever one believes  $p$  one knows oneself to believe  $p$ . But now consider this scenario: you believe that  $p$  and your belief is consciously accessible to you (whatever that turns out to mean). A third party (your spouse,

perhaps, or a therapist) who is taken by you to be an authority on such matters tries to convince you that you do not in fact believe  $p$ .<sup>23</sup> As a result, either you no longer believe that you believe that  $p$ , or your justification for believing that you believe  $p$  is defeated. Either way, you believe that  $p$  but don't know that you believe that  $p$ .<sup>24</sup>

The implication of these examples is that believing  $p$  and knowing oneself to believe  $p$  are two cognitive states. If they were aspects of one state it's hard to see how you could believe that  $p$  without knowing, even tacitly, that you believe that  $p$ . Even if your first-order belief is conscious, you can fail to know that you have it. More controversially, you can also fail to *believe* that you have it, in which case believing that  $p$  and believing oneself to believe that  $p$  are distinct psychological conditions, and it is indeed reasonable to suppose 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. It isn't an irreducible fact about believing that if you believe that  $p$  you know, at least tacitly, that you believe then  $p$ . To this extent, it does need to be explained how we acquire doxastic self-knowledge. MIM supplies an explanation that seems to work, at least in MUM cases, and it is a better explanation than the one offered by reflectivism. Whether it is the best explanation remains to be seen.

#### 4

So far in this paper I have concentrated on IQ cases, where it is explicitly in question whether you have a given belief. I've distinguished two types of IQ case (PEB and MUM) and suggested that Evans' procedure, which I've interpreted along inferentialist lines, is much better suited to answering the question 'Do you believe that  $p$ ?' in MUM cases than in PEB cases. However, there is also the possibility of knowing, or of coming to know, one's own beliefs when it is not in question what one believes. Can the Mediated Inference Model account for our self-knowledge in non-IQ cases? I want to suggest that it can, though the version of MIM that can account for our self-knowledge in such cases is not Evans' version.

It's always helpful when discussing such matters to have a concrete example in mind so here is one: imagine I'm reading *Capital in the Twenty-First Century* by Thomas Piketty. I've heard a lot about the book and am curious to know what it says, though I'm not reading it in order to answer any specific question, other than 'What does Piketty think?'. A few pages into the book I encounter this sentence: 'Economists are all too often preoccupied with petty mathematical problems of interest only to themselves' (Piketty 2014: 32). As I read the sentence I realize that what Piketty is saying here is exactly what *I* think and perhaps have always thought since my days studying economics at university. This is self-knowledge. It is knowledge of what I do and have believed, but it is not self-knowledge acquired by putting into operation my procedure for answering the question 'Are economists all too often preoccupied with petty mathematical problems of interest only to themselves?'. Nor is it self-knowledge that arises in response to the question 'Do you *think* that economists are too often preoccupied with petty mathematical problems of interest only to themselves?'. Nobody has posed this question, which is why this is a non-IQ case. My realization that I think what Piketty thinks is spontaneous, and triggered by reading something I find compelling rather than by reflection on my own beliefs. A person *might* read Piketty in order to figure out what they think about economists but that's not why *I* am reading him.

I've talked about 'realizing' that I think what Piketty thinks, and I take it that realizing that I think that *p* is a way of knowing that I think that *p*. So the obvious next question is: *how* do I realize that I think that economists are all too often preoccupied with petty mathematical problems of interest only to themselves? It might be the case that my realization is 'triggered' by my reading of Piketty but this is hardly an *explanation* of my self-knowledge. If what we are after is an explanation then it is helpful to distinguish three scenarios. In describing these scenarios I'm going to use '*p*' as shorthand for 'economists all too often preoccupied with



petty mathematical problems of interest only to themselves'. I'd like to suggest that the best explanation of my self-knowledge in all three scenarios is in line with MIM.

Here are the three scenarios:

1. *Scenario 1*: as I encounter Piketty's opinion of economists in *Capital in the Twenty-First Century* I find myself in total agreement. Not only am I conscious of agreeing with Piketty that  $p$ , I'm also conscious of having already thought that  $p$ . Consciously agreeing with Piketty that  $p$  is good evidence that I believe  $p$ , and I realize on the basis of this evidence that I believe that  $p$ . Notice that although consciously agreeing with someone that  $p$  is good evidence that I believe  $p$ , it is not the same as believing that  $p$  and doesn't entail that I believe  $p$ : listening to a particularly charismatic speaker I might feel that I agree with them even though, in the cold light of day, I don't share their beliefs. Similarly, I might be conscious of always having thought that  $p$  even though  $p$  isn't what I have always thought. Still, unless there is something wrong with me, what I am conscious of having thought, or of now agreeing with, is a more or less reliable guide to my past and present beliefs. As long as I grasp the connection between agreeing with someone that  $p$  and believing that  $p$  I can justifiably believe and know that I believe that  $p$  on the basis of my awareness of agreeing with Piketty that  $p$ .
2. *Scenario 2*: in this scenario I have never previously considered whether  $p$ . Until now the question whether  $p$  has never crossed my mind, but the moment I read Piketty I am convinced. His statement that economists are too often preoccupied with petty mathematical problems of interest only to themselves produces in me a feeling of conviction, the conviction that  $p$ . Perhaps, as I read Piketty, I think or say to myself 'Of course', but my agreement is not experienced by me as the reaffirmation of something I have always believed. It feels like an indication of a

new belief. Furthermore, the moment I am conscious of being *convinced* that *p* I know, or am able to know, that I *believe p*. My feeling of conviction doesn't entail that I believe *p* but it is good evidence that I believe *p*.<sup>25</sup> I can know on the basis of this evidence that I believe *p* as long as I understand the link between *feeling convinced* that *p* and *believing* that *p*.

3. *Scenario 3*: in this scenario I have never explicitly thought that *p* until now but I've *felt* that *p* at some level ever since I studied economics at university. One way of putting this would be to say that I have *inchoately* believed *p*, and that reading Piketty brings my inchoate belief to light. It enables me to recognize a hitherto inarticulate belief or "proto-belief", and it does this by presenting me with a compelling formulation of what I have always felt. My recognition is grounded in the feelings that Piketty's text produces in me, a feeling of conviction and a sense of familiarity. My agreement with Piketty isn't experienced by me as indicating a new belief *or* as the reaffirmation of an existing belief. It is an in-between case.

My self-knowledge in each of these scenarios is easily and convincingly accounted for by MIM. In each case, I have (psychological) evidence that I believe that *p*, and as long as I understand what this evidence is evidence *for* I can know on this basis that I believe that *p*. The psychological evidence in each case (consciously agreeing that *p*, feeling convinced that *p*) is an occurrent conscious propositional attitude, and the resulting knowledge that I believe that *p* is consciously based self-knowledge.<sup>26</sup> Regardless of whether I *arrive at* knowledge of my belief by inference from my evidence, my *justification* for believing that I believe *p* comes in part from my justification for believing other propositions – propositions about my states of consciousness and what they reveal about my beliefs. I know I believe *p* because I have evidence that I believe it, and my evidence in each case is a state of consciousness.

Belief is standing attitude rather than a state of consciousness but my states of consciousness can disclose what I believe.<sup>27</sup>

How does this account of self-knowledge in non-IQ cases compare with what Evans says about self-knowledge in IQ cases? Looked at in one way the accounts are very different. I don't know that I agree with Piketty by asking myself whether I agree with him; it's *evident* to me that I agree with him without having to execute Evans' procedure, and what makes this evident to me is how I am conscious of reacting to Piketty as I read him. On the other hand, it is arguable that looked at another way my account of self-knowledge in non-IQ cases not fundamentally different from Evans' account of self-knowledge in IQ cases. More cautiously, the two accounts aren't fundamentally different on the assumption that Evans's theory is a version of MIM. The basic idea is the same: regardless of whether I or anyone else has posed the question whether I believe that  $p$ , my knowledge that I believe it is mediated by my knowledge of other mental acts or occurrences (judgements, feelings of conviction) and an implicit understanding of the relationship between these mediating acts or occurrences and my beliefs. What I do not have, either in Evans' original example or in my three scenarios, is epistemologically unmediated self-knowledge.

5

I said at the beginning that I would be arguing for the following claims:

- (A) While Evans identifies a possible route to knowledge of our own beliefs, it's not clear either that we must acquire knowledge of our own beliefs by executing the procedure he describes or that we do generally come to know our own beliefs in this way.
- (B) The self-knowledge that Evans describes is indirect self-knowledge. He describes a procedure for answering questions about what one believes, and it's hard to make sense of this procedure other than on the assumption that the

knowledge it produces is neither epistemologically nor psychologically immediate.

I have said a lot in defence of (B) but what about (A)? My case for (A) has been implicit in the discussion of the last two sections but there is no harm in making it explicit. In effect, there are three ways of reading Evans' theory: as an account of how we *do*, how we *can*, or how we *must* come to know our own beliefs. The sense in which Evans identifies a possible route to self-knowledge is that it is indeed possible for me to know by applying his procedure whether I think there is going to be a third world. The issue is not *whether* this is possible but *how* it is possible. I've suggested that Evans' account is incomplete and that the explanatory gap at its centre can be filled in by reference to MIM. MIM is not opposed to the idea that you can answer the question whether you believe *p* by asking whether *p*. What it says is that it is what you *judge* in response to the latter question that tells you whether you believe *p*.

The problem with saying that it is *necessary* to use Evans' procedure in order to know one's own beliefs is that in non-IQ cases I can know that I believe *p* without asking myself, or being asked, whether I believe *p*. I can also know in some IQ cases whether I believe that *p* without addressing the question whether *p*. These are cases of pre-existing belief. Only in a limited range of IQ cases does Evans' procedure look appropriate, namely, cases in which the question 'Do you believe that *p*?' is read as an invitation to make up my mind whether *p*. However, even in this limited domain I haven't seriously considered whether there might be alternatives to Evans' procedure. We should be cautious about claiming that there is no alternative since this concession might simply reflect a lack of philosophical imagination, or information about the range of ways in which human beings actually arrive at self-knowledge in MUM cases.

This brings us to the question whether Evans' account works as an account of how we actually know our own beliefs. In one sense the answer to this question is clearly 'no' since Evans isn't trying to account to knowledge of our own beliefs in non-IQ cases. His account is limited in scope and it is easy to imagine scenarios in which self-knowledge is acquired by other means. Even if we ignore non-IQ and PEB cases, it's worth pointing out that Evans doesn't supply any empirical evidence that when faced with the question 'Do you believe that  $p$ ?' in a MUM case what we actually do is apply his procedure. For that matter, I also haven't supplied any empirical evidence in support of MIM. All I've done is to put MIM on the table and suggest that it tells a plausible story about how we know our own beliefs in a wide range of cases.

It might seem that the obvious next step would be to look to empirical psychology for evidence in support of MIM, evidence that backs the suggestion that MIM gives an accurate description of how humans beings actually come to know their own minds. But this is only the obvious next step to the extent that MIM is understood as a psychological theory of self-knowledge, a theory of the mental procedures by which we come to know our own beliefs. However, MIM is also intended as a contribution to epistemology, as an account of the justificational structure of our beliefs about own beliefs.<sup>28</sup> The occurrence of judgements, feelings of conviction and experiences of agreeing with another person can be established by attending to our own mental lives. The philosophical question concerns the epistemological significance of such occurrences, and this is the question to which, I contend, MIM supplies a promising answer. The take-home message is this: contrary to what many philosophers think, knowledge of our own beliefs is not based on no evidence. We often know our own beliefs because we are conscious of other mental occurrences that reveal what we believe.

## REFERENCES

- Boyle, M. (2011), 'Transparent Self-Knowledge', *Aristotelian Society*, supplementary volume 85: 223-41.
- Byrne, A. (2011), 'Transparency, Belief, Intention', *Aristotelian Society*, supplementary volume 85: 201-21.
- Carruthers, P. (1996), 'Simulation and self-knowledge: a defence of theory-theory', in P. Carruthers and P. K. Smith (eds.), *Theories of theories of mind* (Cambridge: Cambridge University Press): 22-38.
- Carruthers, P. (2009), 'How we know our own minds: The relationship between mindreading and metacognition', *Behavioral and Brain Sciences*, 32: 1-18.
- Carruthers, P. (2011), *The Opacity of Mind: An Integrative Theory of Self-Knowledge* (Oxford: Oxford University Press).
- Cassam, Q. (2010), 'Judging, Believing and Thinking', *Philosophical Issues*, 20: 80-95.
- Cassam, Q. (2014). *Self-Knowledge for Humans* (Oxford: Oxford University Press).
- Cassam, Q. (2015), 'What Asymmetry? Knowledge of Self, Knowledge of Others, and the Inferentialist Challenge', *Synthese*.
- Evans, G. (1982), *The Varieties of Reference*, ed. J. McDowell (Oxford: Oxford University Press).
- Hieronimi, P. (2009), 'Two Kinds of Agency', in L. O'Brien and M. Soteriou (eds.), *Mental Actions* (Oxford: Oxford University Press):138-62.
- Lawlor, K. (2009), 'Knowing What One Wants', *Philosophy and Phenomenological Research*, 79: 47-75.
- Martin, M. G. F. (1998), 'An Eye Directed Outward', in C. Wright, B. Smith and C. Macdonald (eds.), Knowing Our Own Minds (Oxford: Oxford University Press): 99-121.
- 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. (2004), 'Replies to Heal, Reginster, Wilson, and Lear', *Philosophy and Phenomenological Research*, 69: 455-72.
- Moran, R. (2012), 'Self-Knowledge, "Transparency", and the Forms of Activity', in D. Smithies and D. Stoljar (eds.), *Introspection and Consciousness* ((Oxford: Oxford University Press): 211-36.
- Peacocke, C. (1998), 'Conscious Attitudes, Attention, and Self-Knowledge', in C. Wright, B. Smith and C. Macdonald (eds.), *Knowing Our Own Minds* (Oxford: Oxford University Press): 63-98.
- Piketty, T. (2014), *Capital in the Twenty-First Century*, trans. Arthur Goldhammer (Cambridge, Mass.: Harvard University Press).
- Pryor, J. (2005), 'There is Immediate Justification', in M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology* (Oxford: Blackwell): 181-202.
- Reed, B. (2010), 'Self-Knowledge and Rationality', *Philosophy and Phenomenological Research*, 80: 164-181.
- Scanlon, T. (1998), *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press).
- Schwitzgebel, E. (2011), 'Knowing Your Own Beliefs', *Canadian Journal of Philosophy*, supp. vol. 35: 41-62.
- Shah, N., and Velleman, J. (2005), 'Doxastic Deliberation', *Philosophical Review*, 114: 497-534.

---

<sup>1</sup> This is also a limitation in Richard Moran's Evans-inspired account of self-knowledge. Moran repeatedly emphasizes that he is giving an account of how to answer the question what my belief about something is 'when the question arises'(2004: 467).

<sup>2</sup> Compare the account of 'immediate' justification in Pryor 2005, a discussion to which I am indebted.

<sup>3</sup> In Richard Moran's terminology the question whether *p* is 'outward-directed' whereas the question whether I believe *p* is 'inward-directed' (Moran 2004: 457).

<sup>4</sup> For a defence of inferentialism about self-knowledge see Lawlor 2009, Carruthers 2011 and Cassam 2014.

<sup>5</sup> See Moran 2001: 61 on the significance of the fact that the two questions are different.

<sup>6</sup> Shah and Velleman observe that 'the question "Do I believe that *p*" can mean either "Do I already believe that *p* (that is, antecedently to considering this question?" or "Do I now believe that *p* (That is, now that I am answering the question)?"' (2005: 506).

<sup>7</sup> As Baron Reed notes. See his example of Penny the economist who, in answering the question whether she believes *p*, 'ought to defer to her earlier judgement whether *p*' (2010: 176-7). Recognizing that she already believes *p* 'may count for Penny, not merely as a *reason* to believe that *p*, but as the *answer* to the question *does she believe that p?*' (2010: 177).

<sup>8</sup> Matthew Boyle objects that 'only a madman could draw such an inference' (2011: 227).

<sup>9</sup> I take it that Boyle has a similar worry in mind when he writes that Byrne's approach 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*' (2011: 231).

<sup>10</sup> Here I follow Shah and Velleman's account of judgement and belief. They write that 'a judgement is a cognitive mental act of affirming a proposition.... It is an act because it involves occurrently presenting a proposition, or putting it forward in the mind; and it is cognitive because it involves presenting the proposition *as true*, or, as we have said, affirming it. A belief, by contrast, is a mental state of representing a proposition as true, a cognitive attitude rather than a cognitive act' (2005: 503).

<sup>11</sup> Cf. Shah and Velleman 2005: 508 and Cassam 2010.

<sup>12</sup> Pamela Hieronymi has a different view of this kind of example. She agrees that 'you might deliberate and come to a conclusion that is at odds with the attitudes you continue to hold'. However, she goes on to argue that if you have settled the question whether *p* by concluding that *p* then 'you will, at least for the moment, incur the commitments associated with believing that *p*, and, therefore, that you do, at least for the moment, believe *p* – perhaps despite the fact that you also continue to believe not *p*' (2009: 143). I take the belief that *p* to be a standing dispositional state rather than a momentary occurrence. The dispositions that are relevant to belief are both behavioural and cognitive (see Schwitzgebel 2011: 43-4). To the extent that the subject in Peacocke's example is not disposed to treat degrees from other countries as being of equal value they don't have believe, even momentarily, that they are of



---

equal value. But this doesn't prevent them from mentally *affirming* that degrees from other countries are of equal value. The mismatch is between their continuing, standing state (which is dispositional) and their mental affirmations (which are not). I see no value in muddying the waters by talking about what they believe 'for the moment' or, even worse, what they believe 'occurrently'.

<sup>13</sup> As Moran puts it, 'if the person were entitled to assume, or in some way even obligated to assume, that his considerations for or against believing P (the outward-directed question) actually determined in this case what his belief concerning P actually is (the inward-directed question) then he would be entitled to answer the question concerning his believing P or not by considerations of the reasons in favor of P' (2004: 457). See also Moran 2003.

<sup>14</sup> So this is a place where there is scope for a determined sceptic to insert his knife.

<sup>15</sup> This is Peacocke's objection to the inferential model. See Peacocke 1998: 71-2.

<sup>16</sup> Following Pryor 2005 I take it that all it means for you to have justification to believe some proposition is that it would be epistemically appropriate for you to believe it.

<sup>17</sup> This could be something you say to yourself in inner speech. I agree with Carruthers that 'our conscious occurrent judgements may mostly consist in deployments of imaged sentences, generally the very same sentences that one would use to express those judgements aloud' (1996: 28). I also agree with him that 'it may be that the first metacognitive access subjects have to the fact that they have a particular belief is via its verbal expression (whether overtly or in inner speech)' and that 'such speech, like all speech, will need to be interpreted to extract its significance' (2009: 5). If interpretive access is inferential access then this is another reason to regard knowledge of our own judgements (knowledge of *what* we judge rather than knowledge *that* we judge) is inferential. For further discussion see Cassam 2014, chapter 12.

<sup>18</sup> Krista Lawlor's account of how you know your own desires is the inspiration for this account of how you know you judge that *p*. See Lawlor 2009. When she talks about 'internal promptings' she means such things as simple sensations, thoughts, imaginings, fantasies, and imaged natural language sentences. Her proposal is that inferences from internal promptings may be the basis on which one comes to know one's own desires. I agree. The more radical inferentialist proposal I am now considering is that inference from internal promptings might even play a role in generating knowledge of one's own *judgements*. See Carruthers 2009 and 2011 for a proposal in the same spirit.

<sup>19</sup> Here I follow Scanlon 1998: 21 and Schwitzgebel 2011: 43-4.

<sup>20</sup> I say much more about all this in Cassam 2015.

<sup>21</sup> This is my label, not Boyle's.

<sup>22</sup> This isn't quite the whole story. Even if I knowingly accept that *p*, this isn't sufficient for me to know that I believe that *p* unless I 'understand that what I accept as true just is what I believe' (2011: 237). Now Boyle's account doesn't look all that different from my inferentialist account according to which knowingly judging that *p* enables you to know that you believe that *p* as long as you understand the link between what you judge and what you

---

believe. There is one important difference. According to MIM, the judgement that  $p$  normally produces the belief that  $p$  but the two (the judgement and the belief) aren't identical. For Boyle, in contrast, evaluating a proposition as true (judging, in my terms) 'is not an act one performs to *produce* a belief in oneself; it is one's belief itself' (2011: 236). Despite this difference, versions of most of my five questions about MIM can be raised about Boyle's account.

<sup>23</sup> This is a variation on M. G. F. Martin's example. See Martin 1998: 107.

<sup>24</sup> It might be objected that if the third party really brings it about that you don't believe that you believe that  $p$  then they also thereby bring it about that you don't believe that  $p$ . For if you believe that  $p$  then you must be disposed to self-ascribe this belief but *ex hypothesi* you no longer have this disposition if you are convinced by the third party. On the other hand, you might still be disposed to act as if  $p$  and use it as a premise in further reasoning. To this extent you might still be said to retain the first-order belief despite losing the second-order belief. Now consider the following variation: this time you take no notice of the third party and continue to believe that you believe that  $p$ . Your second-order belief is in fact correct but your *justification* for having it is defeated by expert testimony that, on this occasion, you ignore. Your belief that  $p$  is consciously accessible to you but the presence of a defeater means that you don't know you have it. You don't know – even tacitly- that you have it because your belief that you have it isn't a justified belief.

<sup>25</sup> This is a case in which, as Martin puts it, 'the guidance that one has as to whether one believes something' goes 'via some subjective feeling of conviction' (1998: 116).

<sup>26</sup> As Peacocke points out, occurrent conscious propositional attitudes share with sensations the property that they 'contribute to what, subjectively, it is like for the person who enjoys them' 1998: 64). There is something that it's like to feel convinced that  $p$  but there isn't something that it is like to believe that  $p$ .

<sup>27</sup> Notice that I don't just *have* psychological evidence that I believe  $p$ , I also *know* my evidence, and my knowledge that I believe that  $p$  is mediated by this knowledge. Why must I know my own evidence? What kind of knowledge is this? Is knowledge of my own conscious states or 'internal promptings' direct? For further discussion of these difficult questions see Cassam 2014, chapter 12.

<sup>28</sup> This way of putting things raises deep questions, which I don't have the space to go into here, about the relationship between epistemology and psychology.