

# JUDGING, BELIEVING AND THINKING

Quassim Cassam

University of Warwick

## 1. Introduction

I want to say about the relationship between judging, believing and thinking, and about the epistemology of judgement, belief and thought. I argue in section 2 that judging is not believing and believing is not judging. Judging, unlike believing, is a mental action. Belief is a mental state. How, then, are believing and judging related? Christopher Peacocke claims that 'to make a judgement is the fundamental way to form a belief' (1998: 88), while Tim Crane makes the closely related claim that 'judgement is the formation of belief' (2001: 104). I express some reservations about these claims. Judging that P normally leads one to believe that P but might fail to do so, and there are fundamental ways of forming beliefs in which judgement is not implicated in the way that Peacocke seems to be suggesting. In section 3, I briefly discuss the relationship between judging and thinking. I take it that judging is a form of thinking but not all thinking is judging. In section 4, I comment on the relationship between believing and thinking. There is a sense in which thinking that P is believing that P and a sense in which it is not. It all depends on whether a statement of the form 'S thinks that P' is understood as a report of a mental state of a mental action.

Having explored the relationship between judging, believing and thinking I focus, in section 5, on the epistemological consequences of my account. One issue is whether, when I know that I think or believe or judge that P, there is a way in which I know, or an answer to the question how I know that I think, believe or judge that P. Another issue is whether

and how our knowledge of our own thoughts, beliefs and thoughts can be immediate, that is, not based on observation, inference or evidence.<sup>1</sup> In section 6, I criticize Richard Moran's account of how knowledge of our own beliefs can be immediate. Finally, in section 7, I propose alternative account based on the idea that immediate knowledge of one's own beliefs results from the operation of a sub-personal monitoring mechanism. On this account, less of our self-knowledge is immediate than is commonly supposed. In addition, the proposed explanation of immediacy implies that when we know our own beliefs other than on the basis of observation, inference or evidence there is a sense in which there is no way in which we know what we believe. Talk of there being a way of knowing, or an answer to the question how we know, is more appropriate in relation to self-knowledge that is not immediate.

## 2. Believing and Judging

Belief is a state rather than an action or process. To say that S believes that P is to report on S's mental state rather than on something that S is literally doing or undergoing. Belief is, in this respect, like knowledge. As Williamson observes, the stative character of knowledge and belief is linguistically marked by the impropriety of progressive tenses: 'S is knowing that P' and 'S is believing that P' are both deviant, and this is just what one would expect if knowledge and belief are states.<sup>2</sup> Coming to believe that P, or forming the belief that P, is certainly an event but believing that P is not an event. There is no such thing as 'occurrently' believing that P.<sup>3</sup>

Judging is a mental action.<sup>4</sup> It is true that 'S is judging that P' sounds a little odd, if not straightforwardly deviant, but that is because the most common use of the verb 'judge' is to pick out the mental state of belief.<sup>5</sup> From the fact that 'S judges that P' sounds more

natural than 'S is judging that P' it does not follow that judging is a mental state. To judge that P is to do something, and to do it for a reason.<sup>6</sup> How, then, are judging and believing related? Is it plausible, for example, that to judge that P is to form the belief that P? It is true that judging that P normally leads one to believe that P but is not guaranteed to do so. The following example of Peacocke's makes this vivid:

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 (1998: 90).

So it is false that judgement is the formation of belief if the implication is that whenever S judges that P, S comes to believe that P.

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 cognitive attitude.<sup>7</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>8</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 countries other than my own are of an equal standard to my own and yet my attitude towards this proposition is not the attitude of belief, as evidenced by my hiring decisions and letters of recommendation.

What about the suggestion that, while it is possible to judge that P without believing that P, to make a judgement is nevertheless the fundamental way to form a belief? It is not obvious what makes a way of forming a belief 'fundamental' but let us consider two very common pathways to belief: perception and testimony. I believe that my laptop is in front of me because I can see it, and I believe that this paper is due at the end of the month because I have been told that it is. What is the role of judgement in the formation of these beliefs? A natural thought is that seeing the laptop in front of me can and usually does lead directly to the belief that there is a laptop there. I might be disposed to judge that there is a laptop in front of me because I believe there is a laptop in front of me but the judgement does not mediate the formation of the belief. The same goes for the formation of testimonial beliefs. You tell me that the paper is due at the end of the month and I believe, as a result, that the paper is due at the end of the month. I do not believe that the paper is due at the end of the month as a result of judging that it is due at the end of the month, or judging anything else. Even if the formation of perceptual or testimonial beliefs is somehow sustained by various background beliefs about the reliability of one's senses or of one's informants, this has little to do with the idea that perceptual or testimonial belief-formation is mediated by judgement.<sup>9</sup>

We often form beliefs other than on the basis of perception or testimony. For example, reasoning can lead to the formation of belief, and it is in this context that it seems appropriate to speak of judgement as the fundamental way to form a belief. Suppose that I am presented with a sound and valid argument for some proposition P. I go through the argument and conclude that P. As a result of concluding that P I come to believe that P. Concluding that P is just judging that P, so here we have a case in which the formation of

the belief is that is mediated by judgement. Indeed, it is difficult to see how, at least in normal circumstances, reasoning in favour of P could lead to the belief that P without going via the judgement that P. It is in this sense that, as Shah and Velleman put it, 'the reasoning that is meant to issue or not issue in a belief is meant to do so by first issuing or not issuing in a judgement' (2005: 503).

For this to support the idea that to make a judgement is the fundamental way to form a belief one would need to think that reasoning one's way to a conclusion is the most fundamental way to form a belief. Since the formation of perceptual and testimonial beliefs is not a matter of reasoning one's way to a conclusion, this implies that belief-formation on the basis of perception or testimony is somehow less fundamental than belief-formation on the basis of reasoning. It is not clear why one should accept this implication. So it seems that, in the absence of further argument, we should view with scepticism the suggestion that to make a judgement is the fundamental way to form a belief. There are pathways to belief that do not involve judgement in anything like the way in which belief-formation by reasoning involves judgement, and that are no less fundamental than belief-formation by reasoning.

Even if it is false that to make a judgement is the fundamental way to form a belief, or that judging that P always leads to the belief that P, it is nevertheless plausible that judging that P normally one leads to believe that P. The formation of the belief that P in response to the judgement that P might be blocked or stymied by a prejudice or phobia but this can hardly be the normal case. It is hard to conceive of a subject who judges that P but never, or hardly ever, comes to believe that P. As noted above, judging that P is the act of occurrently putting P forward in one's mind as true. Other things being equal, one would

expect someone who genuinely puts a 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 judging that P does not always lead one to believe that P. 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.<sup>10</sup>

Suppose, then, that judging that P normally leads the judger to believe that P if she does not already believe that P. Must someone who believes that P judge that P? Clearly not. One's words and actions may make it apparent that one believes that P even though one never judges that P. When I am asleep I still believe that  $2+2=4$  but I do not judge that  $2+2=4$ . Perhaps, in that case, it might be suggested that someone who believes that P must at least be disposed to judge that P. This has some plausibility, but is still not obviously correct. One can imagine someone who finds it psychologically impossible mentally to affirm to herself that P but who nevertheless believes that P. She has no disposition to judge that P, even when explicitly asked whether P, but she does in fact believe that P. If some non-human animals are capable of belief but not judgement then that would be another reason not to regard the belief that P as a disposition to judge that P, or as requiring the disposition to judge that P.

To sum up, we have seen that while belief and judgement are closely related, they are by no means identical. For a start, they belong in different ontological categories. In addition, it is neither the case that 'S believes that P' entails 'S judges that P', nor that 'S judges that P' entails 'S believes that P'. One can believe that P without judging that P or, more controversially, without even being disposed to judge that P, and one can judge that P

without believing or coming to believe that P. Judging that P normally leads to the formation of the belief that P but there are fundamental pathways to belief that do not go via judgement. With these elementary reflections in mind, let us now briefly consider the relationship between judging that P and thinking that P.

### 3. Judging and Thinking

If one judges that P does it follow that one thinks that P? There is a use of 'thinks that P' in which to report that S thinks that P is just to report that S believes that P. So if 'S judges that P' does not entail 'S believes that P' then it is equally true 'S judges that P' does not entail 'S believes that P'. In another sense, thinking is any sort of conscious mental activity. This is what Descartes understands by 'thinking'. In this sense, if one consciously judges or mentally affirms that P then, in the very act of judging or mentally affirming that P, one also counts as thinking that P. All judging, on this view, is thinking, and there is no obvious reason to reject this view.

A more challenging question is whether all thinking is judging. Someone who thinks so is Kant. Consider the following passage from his Prolegomena to Any Future Metaphysics:

The sum of the matter is this: the business of the senses is to intuit, that of the understanding is to think. But thinking is uniting representations in a consciousness. This unification originates either merely relative to the subject and is contingent and subjective, or it happens absolutely and is necessary or objective. The uniting of representations in a consciousness is judgement. Thinking is therefore the same as judging, or referring representations to judgements in general (1977: 304).

Thinking, for Kant, is the same as judging because to think and to judge is to do the very same thing, to unite representations in one's consciousness. The key difference for Kant is not between thinking and judging but between subjective thought or judgement and objective thought or judgement, that is, judgements that concern the state of the judger and those that concern mind-independent reality.

On the conception of thinking and judging which I have been assuming, thinking cannot be the same as judging. Judging that P involves mentally affirming that P. To judge that P is to put P forward in one's mind as true but not all thinking involves putting a proposition forward as true. For example, someone who wonders whether P is thinking – thinking about P, as we would say - but is not mentally affirming P. Someone who knows that P is false but imagining what it would be like if P were true is not putting P forward in her mind as true. These and many other such examples show that thinking is not limited to judging, even if all judging is thinking.

This also has a bearing on Kant's claim in the Critique of Pure Reason that the 'only use' which the understanding can make use of concepts 'is to judge by means of them' (A68/ B93). All thinking involves the use of concepts but not all thinking is judging so it seems that what Kant says in this passage must be false. What is somewhat more plausible is that a possessor of concepts must be capable of exercising them in judgements. It follows that all thinking is tied to the capacity to judge, even if not all thinking is judging. To think is to exercise concepts and the concepts one exercises in thinking, even in thinking that does not itself amount to judging, are ones which one must be capable of exercising in judgements. To this extent, we might say that thinking presupposes judging but we should still stop short of saying that thinking is judging.



#### 4. Thinking and Believing

In one sense, someone who thinks that P might be said to believe that P. Equally, someone who believes that P might be said to think that P. If S thinks that it is raining then she presumably believes that it is raining, and if she believes that it is raining then she presumably thinks that it is raining. It might also be held, not uncontroversially, that the linguistic impropriety of 'S is thinking that it is raining' matches that of 'S is believing that it is raining'. All of this might encourage one to draw the conclusion that thinking and believing are the same thing, and hence that both are mental attitudes or states rather than mental acts: 'S thinks that P' seems no less stative than 'S believes that P'.

There is something right about this but also something that is profoundly mistaken. What is right about it is that there is a use of 'thinks that P' in which it is synonymous with 'S believes that P'. To think that P is, indeed, to be in a mental state which consists in the adoption an attitude towards P. On the other hand, it is clearly false that the verb 'think' can only properly used to report of presence of a mental attitude. For example, there is such a thing as thinking about P. Someone who is thinking about P need not believe that P and might never believe that P. 'S is thinking that P' might be deviant but 'S is thinking about P' is not. Someone who is thinking about P is engaged in a mental activity, and the activity in question is not a belief.

On this account, to think is to do something, and there are many different ways of doing something that counts as thinking. Calculating, judging, speculating, wondering are all ways of thinking. They are mental actions which might result or culminate in belief but the beliefs in which they culminate are state rather than actions. As I calculate the sum of  $68+57$  I am thinking. When, as a result of working it out, I come to believe that the sum of

68+57=125 I am not thinking that the sum of 68+57=125, even though it is undoubtedly true that I now think that the sum of 68+57=125. Thinking is not believing and believing is not thinking.

### 5. Epistemological Issues

With these elementary reflections in mind, let us now consider the following epistemological issues: suppose that I believe that P and know that I believe that P. How do I know that I believe that P? If I judge that P and know that I judge that P how do I know that I judge that P? Finally, if I am actively thinking about P how do I know that this is what I am doing? These questions presuppose that there is such a thing as knowing what one believes, judges or thinks, and this might be questioned. Consider, in this connection, Davidson's claim that 'what sets our knowledge of our own minds apart from other forms of knowledge is that there is no answer to the question how we know what we think' (2001a: 301). One reason is that when we know what we think 'there is no way we know' (Davidson 2001b: 66). For if there is a way we know what we think then it would presumably be possible to answer the question how we know what we think by specifying that way.

The denial that there is a way in which we know what we think or believe makes it hard to see how it can be appropriate to speak of knowledge of what we think or believe. We take it, in general, that if S knows that P then there must be a way in which S knows, regardless of whether S knows what that way is. We take it that there must be an answer to the question how S knows that P even if S does not know the answer. So if it is true, as Davidson claims, that there is no way in which we know what we think and no answer to the question how we know then shouldn't we conclude that so-called "self-knowledge" isn't really knowledge?

Before arriving at this conclusion it would be worth considering whether Davidson's claims about the peculiarities of self-knowledge are correct. Take the claim that there is no way in which we know what we believe. It is surely possible to discover that one believes that P by observing one's own behaviour, and this looks like a way of knowing that one believes that P. The problem with this proposal is that it ignores what Davidson and many others take to be a key feature of the epistemology of self-knowledge. It is widely assumed that knowledge of our own minds is normally immediate, that is, not based on inference, observation or evidence. As Davidson remarks, 'it is seldom the case that I need or appeal to evidence or observation in order to find out what I believe; normally I know what I think before I speak or act' (1994: 43). So the challenge is not just to identify a way of knowing our own minds but to identify a way of knowing our own minds which does justice to the fact that self-knowledge normally is immediate.

One response to this challenge would be to appeal to the notion of action-awareness. This is how Peacocke puts it:

The distinctive way in which a subject comes to know of his own mental actions is by taking an apparent action awareness at face value. You judge that it will rain. When so judging, you have an apparent action awareness of your judging that it will rain. It seems to you that you are judging that it will rain. By taking this awareness at face value, you come to know that you judge that it will rain (2009: 193).

Given that thinking is mental action, a similar account can be given of your knowledge that you are, say, thinking about P. When so thinking you have an apparent action awareness of your thinking about P. By taking this awareness at face value you come to know that you are thinking about P. This is your way of knowing that you are thinking that P and how you

know that you are thinking about P. Yet coming to know on the basis of action awareness that you are thinking or judging is not coming to know that you are thinking or judging on the basis of observation, inference or evidence. So there is no conflict with the idea that self-knowledge is normally immediate. Indeed, it might be thought that the action-awareness account explains the immediacy of self-knowledge while keeping hold of the idea that there is a way of knowing our own minds.

The action-awareness account of self-knowledge raises some difficult questions. One is whether self-knowledge based on action awareness is genuinely non-observational and not based on evidence. More would need to be said about the notions of observation, evidence, and action awareness in order to make progress with this question. For present purposes, however, we can let this pass. For even if action-awareness accounts for our knowledge of our own mental actions it does not account for our knowledge of our own beliefs. The question is: how do we know what we believe, given that beliefs are not mental actions? To put it another way, how do we know what we think, when knowing what we think is understood as knowing what we believe as distinct from knowing our own mental doings?

#### 6. Moran on Self-Knowledge

Richard Moran's proposal is that when the question arises what my belief about something is, avowal is a way of answering the question and hence a way of coming to know it. This knowledge is not based on inference, evidence or observation so the idea that we can know our thoughts or beliefs by avowing them accounts for the immediacy of much of our self-knowledge. Avowal is, in turn, to be understood as a way of answering a

question about one's beliefs that obeys the 'Transparency Condition'. This condition states that:

Ordinarily, if a person asks himself the question "Do I believe that P?", he will treat this much as he would a corresponding question that does not refer to him at all, namely, the question "Is P true?" (2001: 60).

Moran's way of putting this is to say that the question "Do I believe that P?" is transparent to the question "Is P true?". Specifically, '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).

The conclusion of my reflection on the reasons in favour of a proposition is a judgement. Suppose, then, that consideration of the reasons in favour of P leads me to judge that P. As we have seen, one might judge that P without believing that P but let us consider the normal case in which judging that P does leads to the formation of the belief that P. In this case, the belief is formed via the judgement that P. This is consistent with the idea that 'the goal of deliberation, whether theoretical or practical, is conviction' (Moran 2001: 131). The point is rather that, when all goes well, theoretical deliberation results in the conviction that P via the mental act of affirming that P in response to the reasons in favour of P.

What is the epistemological significance of this account of the relationship between judging and believing? Suppose that my recognition of the reasons in favour of P leads me to judge that P, and that I know that I judge that P. And suppose, also, that my knowledge that I judge that P is taken for granted. Then the next question is: what is the relationship between knowing that I judge that P and knowing that I believe that P? If judging that P

were equivalent to believing that P then knowing that I judge that P would amount to knowing that I believe that P. But it is false that judging that P is believing that P since, as we have seen, it is possible to judge that P without believing or coming to believe that P. Nevertheless, one might still think that knowing that I judge that P amounts to knowing that I believe that P in at least the following sense: if I know that I judge that P, and am entitled to assume that my judgements normally determine my beliefs, then I can know or conclude that I believe that P.

Something along these lines is suggested by Moran's response to what he regards as a major challenge facing his theory. The challenge is to explain what right I have to think that 'my reflection on the reasons in favour of P (which is one subject-matter) has anything to do with the question of what my actual belief about P is (which is a quite different subject matter)' (2003: 405). He responds that I would have a right to assume that this is the case 'if I could assume that what my belief here is was something determined by the conclusion of my reflection on those reasons' (ibid.). He adds:

And now, let's ask, don't I make just this assumption, whenever I'm in the process of thinking my way to a conclusion about some matter? I don't normally think that my assessment of the reasons in favour of P might have nothing to do with what my actual belief is, and it's hard to imagine what my thinking would be like if I did normally take this to be an open question (2003: 405-6).

Since the conclusion of my reflection on the reasons in favour of P is a judgement, the assumption that my belief concerning P is determined by the conclusion of my reflection on the reasons in favour of P is the assumption that, in the case in which I end up believing that P as a result of deliberation, my believing that P is determined by my judging that P. It

now appears that the role of this assumption in Moran's account is to facilitate the transition from knowledge of what one judges to knowledge of what one believes. If I know that I judge that P then I am in a position to know that I believe that P, but only if I take what I believe to be determined by what I judge.

If this is the proposal then it is not clear how it accounts for the immediacy of our knowledge of our own beliefs. My knowledge that P is immediate if and only if it is not based on observation, evidence, or inference of any kind but if I know that I believe that P on the basis of my knowledge that I judge P then there is a sense in which my knowledge of my belief based on evidence.<sup>11</sup> The case for thinking that this is evidence-based knowledge is this: my judging that P is neither identical with nor entails that I believe that P. However, my judging that P normally leads (in the case in which I don't already believe that P) to my forming the belief that P, so the fact that I judge that P raises the probability that I believe that P. It makes it likely that I believe that P and is, in this sense, a reliable sign that I believe that P. But this is just what it is for one thing to be evidence for another.<sup>12</sup> Furthermore, it is not just that my judging that P is evidence that I believe that P. It is also evidence I have, to the extent that I know that I judge that P and am aware of the connection between what I judge to be the case and what I believe. Finally, my knowledge that I believe that P is based on evidence in my possession in the following sense: I know I believe that P because I know that I judge that P, and would not know in this case that I believe that P if I did not know that I have just judged or concluded that P in response to the reasons in favour of P. When this is the basis on which I know that I believe that P my knowledge is based on evidence and so not epistemically immediate.

One option at this point would be to question the intuition that self-knowledge is normally immediate. The thought would be that if avowal is not a source of immediate self-knowledge then so much the worse for the immediacy intuition, especially when account is also taken of the haziness of the distinction between knowledge that is, and knowledge that is not, 'based on evidence'. But before abandoning the immediacy intuition it would be worth considering an alternative to Moran's account of self-knowledge.

### 7. Sub-Personal Monitoring and Immediate Self-Knowledge

As Shah and Velleman point out, 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 considering the question?"' (2005: 506). Even if the latter question can be answered by considering the reasons in favour of P and forming a belief with respect to this proposition 'one cannot answer the question whether I already believe that P in a way that begins with forming the belief' (ibid.). This is the problem of antecedent belief. In essence, the problem is that it does not appear to be possible to come to know one's antecedent beliefs by now avowing them.<sup>13</sup>

Here is a relatively straightforward solution to the problem of antecedent belief: when I come to believe that P the representation that P enters my belief store or, as is sometimes said, my Belief Box.<sup>14</sup> So when I am asked whether I (already) believe that P, what this question calls for is a search of my Belief Box. If the belief that P is found in my Belief Box this leads to the formation of the second-order belief that I believe that P, and this second-order belief might itself end up in my Belief Box. The three questions that now arise are the following:

- (a) Who or what is responsible for the search of my Belief Box?



(b) What are the circumstances in which I count as knowing and not merely believing that I believe that P?

(c) How does any of this help to explain the immediacy of my knowledge that I believe that P?

With regard to (a), the proposal is that the search of my Belief Box is not carried out by me, the subject, but by one of my sub-personal monitoring mechanisms, that is, 'a distinct mechanism that is specialized for detecting one's own mental states' (Nichols and Stich 2003: 163). The speed and ease with which this mechanism operates explains the speed and ease with which the second order belief is formed. The discovery of the belief that P in my Belief Box leads directly to the formation of the second-order belief. There is no mediation by judgement or anything else in the formation of the second-order belief.

For my second-order belief to count as knowledge it must satisfy the conditions for knowing. Even if one is sceptical about the prospects for a fully reductive analysis of the concept of knowledge it is still plausible that there are non-trivial necessary conditions for knowledge and that these conditions include a safety condition: I count as knowing that I believe that P only if my belief that I believe that P could not easily have been false. Since it is easy to conceive of the monitoring mechanism as giving rise to second-order beliefs that satisfy this and other conditions on knowledge it is easy to conceive of my second-order belief that I believe that P as amounting to knowledge that I believe that P. In this account of self-knowledge all the work is done by one's assumptions about the reliability of the monitoring mechanism.

There is a straightforward sense in which, according to this account, my knowledge that I believe that P is immediate: the sub-personal monitoring of my belief store is not a

form of perception or inference. Since the resulting knowledge is also not based on behavioural evidence the only issue is whether it can be said to be based on evidence of some other kind. But what kind of evidence would that be? In Moran's account, I know that I believe that P on the basis of my knowledge that I judge that P but the present account tells a very different story: it is not being proposed that I am conscious of judging that P and then conclude on this basis that I believe that P. The second-order belief is not mediated by judgement or by consideration of the reasons in favour of P and there is, in this sense, no evidence on the basis of which I know that I believe that P. My second-order belief is produced by a reliable mechanism but this puts no pressure on the idea that the self-knowledge to which it gives expression is not based on inference, observation or evidence.

This explanation of the possibility of immediate self-knowledge draws on aspects of what Nichols and Stich call Monitoring Mechanism (MM) theory of self-awareness.<sup>15</sup> As they point out, a good theory of self-awareness needs to be able to explain the fact that 'when normal adults believe that P, they can quickly and accurately form the belief I believe that P' (2003: 160). In order to implement this ability, 'all that is required is that there be a Monitoring Mechanism (MM) that, when activated, takes the representation P in the Belief Box as input and the representation I believe that P as output' (2003: 160-1). The Monitoring Mechanism simply has to copy representations from the Belief Box and embed copies of them in a schema of the form 'I believe that...'. Stich and Nichols do not draw attention to the consequences of their view for the issue of immediacy, but it seems obvious that if a Monitoring Mechanism is sufficiently reliable to produce knowledge of one's own

beliefs then the knowledge to which it gives rise is both psychologically and epistemically immediate.

We now have a relatively straightforward answer to the question: 'how is immediate self-knowledge possible?'. It is possible to the extent that one's beliefs about one's own beliefs are produced by a reliable Monitoring Mechanism. A creature endowed with such a mechanism would be able to know its own beliefs other than on the basis of inference, observation or evidence, and it is plausible that we are such creatures. So the suggestion is not merely that we now have an account of how a creature could, in principle, come to have immediate knowledge of its own attitudes but also an account of how we actually come to have immediate self-knowledge.

The most pressing of the many questions raised by this account of immediate self-knowledge is whether it can or should accommodate Moran's Transparency Condition. I have represented the MM account as responding to the problem of antecedent belief but, as noted above, there is also the phenomenon of coming to know whether one believes that P by considering the reasons in favour of P. This is the phenomenon of transparency, and it is a genuine phenomenon. It is one thing to dispute the immediacy of the self-knowledge that is made available by the transparency procedure and another to deny that it is possible for us to acquire self-knowledge in the way that Moran describes. Transparency is a phenomenon which any account of self-knowledge should seek to accommodate.

The problem for the MM theorist is that the procedure that Moran describes is one that takes place at the personal rather than the sub-personal level. It is for me rather than my sub-personal Monitoring Mechanisms to consider the reasons in favour of P, and it hardly needs saying that such mechanisms do not make assumptions about the extent to which our

beliefs are determined by reflection on the reasons in favour of them. Talk of mechanisms for detecting one's own beliefs implies that the beliefs in question are already there, but there is also the case in which I do not already believe that P. Instead, I come to believe that P in the course of considering the question whether I have this belief. In this case, there is no pre-existing belief for my Monitoring Mechanism to latch onto, and the MM theorist does not appear to have a story to tell about this pathway to self-knowledge.

At this point, there are two directions in which the discussion can go. The first would be to try to demonstrate that the MM account can make space for an analogue of the Transparency Condition. The idea would be that when the Monitoring Mechanism searches a Belief Box it is somehow sensitive to the specific ways in which the contents of the box are responsive to evidence. On the other hand, perhaps it is simply a mistake to look for a unified theory of self-knowledge, one that seeks to identify a single way in which we come to know our own beliefs. The alternative would be to accept a hybrid theory on which the MM tells part of the story about self-knowledge, the part that is best equipped to solve the problem of antecedent belief. This leaves room for the idea that avowing a belief is also way of coming to know that one has it. The fact remains, however, that avowal in Moran's sense is not a source of immediate self-knowledge. When it comes to explaining how immediate self-knowledge is possible the MM account is in much better shape. If it only accounts for the immediacy of some of our self-knowledge then this only serves to confirm the suspicion that less of our self-knowledge is immediate than is commonly supposed.

Where does this leave the debate over whether there is a way in which we know our own beliefs or an answer to the question how we know what we believe? Much depends on how the notion of a way of knowing is understood. Consider Williamson's claim that 'if one

knows that A, then there is a specific way in which one knows; one can see or remember or... that A' (2000: 34). On this account, knowledge or 'knows' is a determinable, and specific ways of knowing like seeing and remembering are its determinates. Alternatively, ways of knowing might be conceived of as pathways to knowledge or as ways of coming to know.<sup>16</sup> Either way, seeing and remembering are personal-level ways of knowing. It is the subject S who sees or remembers that A, and pointing out that S sees or remembers that A is an answer to the question "How does S know that A?". The answer is one that explains S's knowledge by reference to something that S, in a sense, does.<sup>17</sup>

In contrast, it is not the subject who monitors his or her own Belief Box. When the question arises whether I believe that P, and the answer "yes" pops into my head as a result of the operation of a reliable Monitoring Mechanism, this is an explanation of my knowledge but the explanation is not a personal-level explanation. It does not refer to something I have done in order to know that P, and is not an explanation of the form 'S  $\Phi$ s that he believes that P and thereby knows that he believes that P'. To say that the answer to the question whether I believe that P pops into my head is to imply that there is nothing that I have to do in order to discover that I believe that P, even though it is undoubtedly true that something has to happen - sub-personally- in order for me to know that I believe that P.

If ways of knowing are personal-level pathways to knowledge, or personal-level determinates of the determinable 'knows', then the implication of this discussion is that one can know that one (already) believes that P without there being a way one knows. In contrast, avowing a belief would count as a way of knowing that one has it but, as we seen, there is a question about the immediacy of the resulting knowledge. In general, it is easier to see how self-knowledge can be immediate if its acquisition is due to the operation of a

sub-personal Monitoring Mechanism rather than the assessment of reasons. On the MM account there is still a way of knowing that one believes that P if all this means is that it is not a complete mystery how I know that I believe that P. To say that there must be a way in which I know what I believe, or an answer to the question how I know, is to insist on the explicability of this form of self-knowledge. For someone who understands the MM account, knowledge of one's own beliefs is by no means inexplicable. The explanation is rather different from the personal-level explanation of other forms of knowledge but that is as it should be.<sup>18</sup>

## REFERENCES

- Boyle, M. (2009), 'Two Kinds of Self-Knowledge', Philosophy and Phenomenological Research, LXXVIII.
- Cassam, Q. (2007b), 'Ways of Knowing', Proceedings of the Aristotelian Society, CVII.
- Crane, T. (2001), Elements of Mind (Oxford: Oxford University Press).
- Davidson, D. (1994), 'Knowing One's Own Mind', in Q. Cassam (ed.) Self-Knowledge (Oxford: Oxford University Press).
- Davidson, D. (2001a), 'Comments on Karlovy Vary Papers', in P. Kotatko, P. Pagin and G. Segal (eds.) Interpreting Davidson (Stanford: CSLI Publications).
- Davidson, D. (2001b), 'What is Present to the Mind', in Subjective, Intersubjective, Objective (Oxford: Oxford University Press).
- Kant, I. (1932), Critique of Pure Reason, trans. Norman Kemp Smith (London: Macmillan).
- Kant, I. (1977), Prolegomena to Any Future Metaphysics, original trans. P. Carus, rev. J. Ellington (Indianapolis: Hackett Publishing Company).
- Kelly, T. (2006), 'Evidence', in E. N. Zalta (ed.) The Stanford Encyclopedia, URL: <http://plato.stanford.edu/entries/evidence/>.
- Moran, R. (2001), Authority and Estrangement: An Essay on Self-Knowledge (Princeton and Oxford: Princeton University Press).
- Moran, R. (2003), 'Responses to O'Brien and Shoemaker', European Journal of Philosophy, 11.
- Moran, R. (2004), 'Replies to Heal, Reginster, Wilson, and Lear', Philosophy and Phenomenological Research, LXIX.
- Nichols, S & Stich, S. (2003), Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds (Oxford: Oxford University Press).

- Peacocke, C. (1998), 'Conscious Attitudes, Attention, and Self-Knowledge', in C. Wright, B. Smith and C. Macdonald (eds.) Knowing Our Own Minds (Oxford: Clarendon Press).
- Peacocke, C. (2007), 'Mental Action and Self-Awareness (I)', in B. McLaughlin and J. Cohen (eds.) Contemporary Debates in Philosophy of Mind (Oxford: Blackwell Publishing Ltd.).
- Peacocke, C. (2009), 'Mental Action and Self-Awareness (II)', in L. O'Brien and M. Soteriou (eds.) Mental Actions (Oxford: Oxford University Press).
- Shah, N and Velleman, D. (2005), 'Doxastic Deliberation', The Philosophical Review 114.
- Soteriou, M. (2009), 'Mental Agency, Conscious Thinking, and Phenomenal Character', in L. O'Brien and M. Soteriou (eds.) Mental Actions (Oxford: Oxford University Press).
- Williamson, T. (2000), Knowledge and its Limits (Oxford: Oxford University Press).



---

<sup>1</sup> This is Richard Moran's gloss on what it would be for self-knowledge to be immediate. See Moran 2001: 91. Elsewhere, he only seems to require for immediacy that the knowledge in question not be based on observation or inference. See, for example, Moran 2004: 423. I operate with the more demanding conception of immediacy. The immediacy that is in question here is epistemic rather than psychological.

<sup>2</sup> See Williamson 2000: 35.

<sup>3</sup> As Crane observes, ‘“occurrent belief” is a myth’ (2001: 108).

<sup>4</sup> As Peacocke has frequently emphasized. See, for example, Peacocke 2007.

<sup>5</sup> As Matthew Soteriou also notes in an important paper. See Soteriou 2009: 240.

<sup>6</sup> In Peacocke's words, ‘judgement is a conscious rational activity, done for reasons, where these reasons are answerable to a fundamental goal of judgement: that it aims for truth’ (1998: 88).

<sup>7</sup> Here I closely follow Shah and Velleman's account of judgement and belief. They write that ‘a judgement is a cognitive mental act of affirming a proposition (although.... not all affirmations are judgements. 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>8</sup> Cf. Shah and Velleman 2005: 508.

<sup>9</sup> The point that not all beliefs need be the products of judgements is one that Crane also makes. See Crane 2001: 104.

---

<sup>10</sup> On the idea that judgements are a sub-class of affirmations see Shah and Velleman 2005: 504.

<sup>11</sup> My knowledge in this case also looks inferential but this is obviously not what Moran thinks.

<sup>12</sup> As Williamson remarks, what is required for e to be evidence for the hypothesis h is that 'e should speak in favour of h' and should itself have 'some kind of creditable standing' (2000: 186). In probabilistic terms, e speaks in favour of h if it raises the probability of h. Kelly points out that 'the notion of evidence is that of something which serves as a reliable sign, symptom, or mark of that which it is evidence of' (Kelly 2006).

<sup>13</sup> Shah and Velleman's own response to the problem of antecedent belief is worth quoting: 'If the question is whether I already believe that P, one can assay the relevant state of mind by posing the question whether P, and seeing what one is spontaneously inclined to answer. In this procedure, the question whether P serves as a stimulus applied to oneself for the empirical purpose of eliciting a response. One comes to know what one already thinks by seeing what one says - that is, says in response to the question whether P' (2005: 506). This procedure requires one to refrain from reasoning as to whether P since that reasoning might alter the state of mind one is trying to get at. In addition, testing one's spontaneous response to the question whether P 'may yield good evidence as to whether one already believes that P, but that evidence isn't conclusive: one's first thought upon entertaining a question may be misleading as to one's pre-existing attitude' (2005: 507). On this account, knowledge of what one already believes is clearly based on evidence and therefore not immediate. See Boyle 2009 for some pertinent criticisms of Shah and Velleman's proposal.

---

<sup>14</sup> A Belief Box represents a functionally characterized set of mental states. As Nichols and Stich point out, positing such a box 'does not commit a theorist to the claim that ... the states are spatially localized in the brain, any more than drawing a box in a flow chart for a computer programme commits one to the claim that the operation that the box represents is spatially localized in the computer' (2003: 11).

<sup>15</sup> It also draws on conversations and correspondence with Tim Williamson. It was Williamson who first drew my attention to the possibility of exploiting something like the MM theory to explain the immediacy of self-knowledge. He does not, however, endorse the present approach to self-knowledge.

<sup>16</sup> See Cassam 2007 for more on this conception of a 'way of knowing'.

<sup>17</sup> One might be reluctant to accept the implication that seeing is an action but seeing is related to looking and looking is certainly an action.

<sup>18</sup> I thank Tim Williamson for helpful discussions of the themes of this paper.