

MORAN ON SELF-KNOWLEDGE, AGENCY AND RESPONSIBILITY

CARLOS J. MOYA
Departamento de Metafísica
y Teoría del Conocimiento
Universidad de Valencia
Carlos.Moya@uv.es

SUMMARY: In this paper I deal with Richard Moran's account of self-knowledge in his book *Authority and Estrangement*. After presenting the main lines of his account, I contend that, in spite of its novelty and interest, it may have some shortcomings. Concerning beliefs formed through deliberation, the account would seem to face problems of circularity or regress. And it looks also wanting concerning beliefs not formed in this way. I go on to suggest a diagnosis of these problems, according to which they would arise out of a view of agents too strongly dependent on the will.

KEY WORDS: first person, belief, deliberation, avowals

RESUMEN: Este trabajo se ocupa de la concepción del autoconocimiento en el libro de Richard Moran *Authority and Estrangement*. Tras presentar las líneas maestras de dicha concepción, sostengo que ésta, a pesar de su novedad e interés, podría adolecer de defectos importantes. Así, con respecto a las creencias formadas mediante la deliberación, la propuesta de Moran parece enfrentarse a problemas de circularidad o de regreso. Y parece también insatisfactoria acerca de creencias no formadas de ese modo. Finalmente, sugiero un diagnóstico de estos problemas, según el cual éstos surgirían de una concepción de los agentes excesivamente dependiente de la voluntad.

PALABRAS CLAVE: primera persona, creencia, deliberación, declaraciones

In his admirable book *Authority and Estrangement*, Richard Moran has taken a fresh, and refreshing, stance toward the old problem of self-knowledge. This problem, which Descartes brought to the very centre of modern philosophy, may be described as follows. Judgements about one's own beliefs, intentions, and other mental states are apparently made in an immediate (non-inferential) way, in that they do not rely on the sort of behavioural or other empirical evidence one takes into account in coming to know about other people's minds or about the external world; and nonetheless, in spite of this lack of evidential basis, they are especially authoritative and are taken to bear a special presumption of truth. These judgements are essentially first personal and seem to manifest a peculiar knowledge one has of one's own mental life. These features are intriguing partly because, in most other cases, lack of empirical evidence counts against a judgement

or belief's claims to truth and knowledge. Beyond these puzzling traits, self-knowledge also deserves attentive philosophical research by virtue of its indispensability for such abilities as rational agency and choice, critical evaluation of one's beliefs and reflective self-formation. The problem of self-knowledge is partly the problem of providing a reasonable and non-mysterious explanation of this surprising cognitive achievement.

For the last decades, this problem has been the object of philosophical discussion mainly in its relation to other philosophical doctrines. In the context of epistemology, it has been addressed in connection with such questions as foundationalism and introspection. In the philosophy of mind it has been dealt with in its relationships to the discussion about Cartesian and anti-Cartesian views of mind, to psychoanalysis or to cognitive psychology, as well as to externalism about the content of psychological attitudes.¹ In the philosophy of action, interest in self-knowledge has been driven in turn by the special bearing that the first-person perspective seems to have on our agency and the actual exercise thereof.² Interesting as these debates may be, they have distracted the attention from the question about the nature of self-knowledge itself. Participants in this discussion have frequently taken for granted, without much criticism, particular accounts of self-knowledge. Theoretical accounts, in Moran's terms, have been especially prominent among them. It is an important merit of Moran's that, in his book, he is directly concerned with the nature of self-knowledge itself, a question that would seem to be conceptually prior to any worries about the relationships of self-knowledge with other matters. Another outstanding value of Moran's book is the insistence on the crucial importance of self-knowledge for our mental health and freedom, against the —I would think irresponsible—negation of its existence by some modern authors, who tend to see it as a sort of pre-scientific, obscurantist myth. Let me outline briefly some central traits of Moran's view and then proceed to a critical discussion of it.

¹ The central issue of this last debate could be put as follows: if what we believe, intend or desire is partly determined by features of our external environment, which we can be fully ignorant of, how could we possibly know our beliefs, intentions or desires, which content is constitutive of, in an immediate, non-inferential way, without relying on empirical evidence about environmental facts, and none the less with special authority?

² John Perry's paper "The Problem of the Essential Indexical" has been especially influential in the discussion about this topic.

Moran rejects deflationist approaches to self-knowledge, according to which self-knowledge would not be a substantial epistemological achievement. At the same time, he argues against purely contemplative, theoretical views, either spelled out in the classical, Cartesian-like terms of inward glance or inner perception (introspection), or in the more recent idioms of second-order, reflective attitudes.³ Moran conceives of self-knowledge as an essentially practical ability, constitutively related to our agency, freedom and responsibility. He focuses mainly on beliefs, with some references to other attitudes, such as intentions and desires. Though he does not deny that self-knowledge can take the form of a self-attribution of mental attitudes from a third-person stance, on the basis of behavioural and other evidence about oneself, self-knowledge, in so far as it concerns agency and responsibility, manifests itself in the form of *avowals* of one's attitudes from a committed, endorsing, first-personal stance. At the end of chapter 4 of his book, Moran writes: "In this chapter I have argued the case for seeing the ability to avow one's beliefs as the fundamental form of self-knowledge, one that gives proper place to the immediacy of first-person awareness and the authority with which its claims are delivered" (2001, p. 150). Avowals, in Moran's technical use of the term, are declarations of one's belief that *P* on the basis of the same reasons that would justify one's assent to *P* itself.⁴ This is what Moran calls the "Transparency Condition": "With respect to belief, the claim of transparency is that from within the first-person perspective, I treat the question of my belief about *P* as equivalent to the question of the truth of *P*" (2001, pp. 62–63).

Moran holds that our sense of agency and responsibility with respect to our own beliefs is essentially linked to our ability to avow them in conformity to the Transparency Condition: "[O]nly if I can see my own belief as somehow 'up to me' will it make sense for me to answer a question as to what I believe about something by reflecting exclusively on that very thing, the object of my belief" (2001, pp. 66–67). Moran explicitly rejects doxastic voluntarism, the thesis that we can adopt or reject beliefs at will. Responsibility and agency concerning our own beliefs and other attitudes depend rather

³ He writes: "The special features of first-person awareness cannot be understood by thinking of it purely in terms of epistemic access (whether quasi-perceptual or not) to a special realm to which only one person has entry. Rather, we must think of it in terms of the special responsibilities the person has in virtue of the mental life in question being *his own*" (Moran 2001, p. 32).

⁴ "An avowal is a statement of one's belief which obeys the Transparency Condition" (Moran 2001, p. 101).

on our forming and having them in a way that is *responsive to reasons* for and against them,⁵ thus in a way that respects transparency:

One is an agent with respect to one's attitudes insofar as one orients oneself toward the question of one's beliefs by reflecting on what's true, or orients oneself toward the question of one's desires by reflecting on what's worthwhile or diverting or satisfying [...]. There is a role for the agent here insofar as we may speak of a person's responsibility for his attitudes. (Moran 2001, p. 64)

In so far as an agent conceives of her beliefs or desires from a deliberative stance, that is, from the perspective of the reasons she has for forming or maintaining them, she takes it to be up to her to adopt, suspend or reject them. Concerning desires, Moran writes:

The person's responsibility here is to make his desire answerable to and adjustable in the light of his sense of some good to pursue [...]. At the beginning of his practical reasoning he was not aiming to produce a particular desire in himself (as he might with respect to another person), but rather holding open his desire to how the balance of reasons falls out. (Moran 2001, pp. 118, 119)

Sensitivity and responsiveness to reasons is thus essential to the sort of agency and responsibility involved in self-knowledge, which keeps this account quite far from doxastic voluntarism. First-person authority is not so much an epistemic privilege as a matter of an agent's taking responsibility for her thoughts on the basis of the reasons that justify them.⁶

An attitude that an agent finds and ascribes to herself owing to explanatory, rather than justifying, reasons, on the almost exclusive basis of evidence, behavioural and otherwise, is not an attitude that she has authority and control over. This is especially the case with obsessive thoughts or compulsive desires. These are, so to speak,

⁵ It is worth pointing out that, in John Fischer and Mark Ravizza's theory, responsiveness to reasons is an essential requirement of moral responsibility. Cf. Fischer and Ravizza 1998.

⁶ "[T]he primary thought gaining expression in the idea of 'first-person authority' may not be that the person himself must always 'know best' what he thinks about something, but rather that it is *his business* what he thinks about something" (Moran 2001, p. 123).

opaque to the agent and beyond the reach of her rational authority. Self-ascription of these attitudes does not conform to the Transparency Condition. Immediacy and authorship are absent. On the contrary, in the case of avowals,

it is because the agent's perspective is characterized by the dominance of justifying reasons over explanatory ones that the immediacy of avowal is not only a possibility for the agent but also something of a requirement, in the sense that it is an expression of the person assuming responsibility for his thought and action. (Moran 2001, p. 131)

Moran places himself explicitly within a tradition that establishes a tight connection of reflection with freedom and reason. Moran refers to Kant, Locke and early Stoicism as representatives of this tradition, but devotes special attention to Sartre. In Sartre's case, the idea would be this: in becoming reflectively aware of some attitude or impulse in myself, I thereby become free to endorse, permit or reject it. Self-consciousness makes choice in this respect unavoidable for me. As far as I become self-consciously aware of an impulse, desire or belief of mine, I cannot escape the question of what attitude I shall take towards it, whether of "endorsement, permission or disapproval" (Moran 2001, p. 147). Choice is unavoidable, for remaining inactive with respect to my state of mind is also choosing to allow it to continue as part of my psychic life.

Stuart Hampshire and Christine Korsgaard are also authors that Moran refers to in some detail as belonging in this tradition. It is worth noticing that, in this context, he also refers to Descartes, who has usually been considered as a central example of a theoretical, quasi-perceptual conception of self-knowledge. Moran focuses on Descartes' account of the possibility of error in the *Meditations*, on the basis of a distinction he draws between the intellect, which as such is restricted to passively considering ideas, and the agent's (free) will, whose business is to actively form judgments to the effect that things are thus and so.⁷

In the final chapter of his book, Moran presents and discusses what might be called pathologies and paradoxes of self-knowledge. This part of his work is highly interesting and thought-provoking,

⁷ This allows Descartes to ascribe responsibility for our errors not to God but to ourselves, who freely judge without a clear and distinct intellectual perception of the corresponding subject.

but I will not get into it.⁸ My purpose in this paper is to understand and discuss his view of the nature and possibility of self-knowledge. To this end, the brief presentation I have made so far should be sufficient. Let me now proceed to a more critical approach to Moran's proposal.

Philosophical theories of self-knowledge are usually constructed on the basis of reflection on some central examples of the mental and of some basic insights inspired in them, which are then extended to other cases with the aim of covering and accounting for the whole field. So, for instance, it is generally acknowledged that empiricist theories take conscious experiences, such as sensations, as central examples of the mental, and phenomenological consciousness of them, as central cases of self-knowledge. These approaches find difficulties, however, when they confront so-called intentional or propositional attitudes, such as beliefs, which do not seem to be "felt" or present to our phenomenological consciousness at all. In Moran's case, it is beliefs that are taken as paradigmatic examples of the mental and, concerning self-knowledge, Moran's account rests upon cases of belief formation on the basis of reasons and deliberation. These are cases in which beliefs are avowed as conclusions of the agent's deliberation, in a way that respects the Transparency Condition. However, as I shall try to show, avowals apparently not based on reasons and knowledge of the reasons we make use of in deliberation are not easy to accommodate within the model based on the central cases.⁹

As I see things, two philosophers play a distinguished role in the foundations of Moran's proposal and the intuitions that sustain it, namely Gareth Evans and Elizabeth Anscombe. Concerning the former, the following quotation of *The Varieties of Reference*, in which Evans is attempting to interpret a remark by Wittgenstein, is specially important:

[I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward —upon the world. If someone asks me "Do you think there is going to be a third world war?", I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?" I get myself in a position to answer the question

⁸ Interesting treatments of this problem can be found in O'Brien 2003, Reginger 2004 and especially in Lear 2004.

⁹ Knowledge of intentional attitudes other than belief, especially desires, as well as of sensations and conscious experiences, is also difficult to account for from Moran's perspective, but I shall not get into it owing to space limitations.

whether I believe that p by putting into operation whatever procedure I have for answering the question whether p . (Evans 1982, p. 225; see Moran 2001, p. 61)¹⁰

Moran's notion of avowal as a declaration of belief that respects transparency is closely connected with the insight contained in this text. Suppose that I am asked the question at issue and that I have not considered this matter before. I then proceed to attend to reasons for and against the outbreak of a third world war. Suppose that I find the reasons against stronger than the reasons for and go to the conclusion that such a war is a very unlikely event. Now there is nothing else I must do in order to answer the question about what I believe. My conclusion about the matter determines at the same time my belief about it. And when I respond: "No, I don't think so", this is an avowal (of belief) in Moran's sense of this term. According to him, this avowal is a manifestation of self-knowledge, and indeed of the fundamental form of it (cf. Moran 2001, p. 150).

It is tempting to think that my knowledge of the belief I am avowing is already explained through the description of this deliberation process. However, in a reply to George Wilson (Wilson 2004), Moran denies this: "[E]ven of the beliefs that *are* the upshot of explicit deliberation, my claim is not that this deliberation produces both the belief and the self-awareness of it" (Moran 2004, p. 466). And elsewhere, in a reply to Lucy O'Brien (2003) and Sydney Shoemaker (2003), he justifies this contention as follows: "For of course I may conclude something [about a subject matter] [...] and never bother to reflect on the fact that I have come to this belief. I take it that's the normal case" (Moran 2003, p. 414). But then, since, even in Moran's own lights, an answer to the first-order question whether P is not ipso facto an answer to the second-order question whether I believe that P ,¹¹ an agent's knowledge about this belief (i.e., self-knowledge proper) needs some additional explanation, which goes beyond an appeal to the process of deliberation and transparency. In this context, Moran writes:

What I do try to account for is how it can make sense for the first-person question about one's belief to be answered by way of reflection on the reasons relevant to the truth of the belief, rather than by reflection on

¹⁰ Moran (2001, p. 61) quotes also Rod Edgley, who, before Evans, expresses this same contention as well.

¹¹ I owe this way of putting things to an anonymous referee for *Crítica*.

behavioral or other evidence that would be relevant to the psychological ascription to a particular person. (Moran 2004, p. 468)

This, however, is just a reaffirmation of the transparency of first-person avowals, but does not take us very far for what concerns knowledge of the avowed belief. In order to get some clarity about this crucial issue, I think we should look elsewhere, at another important insight that underlies Moran's proposal. It derives from an interpretation of Anscombe's remarks about the knowledge we have of our intentional actions in her classic essay *Intention* (cf. Anscombe 1963). Anscombe relates the concept of intentional action to the concept of reasons as follows:

What distinguishes actions which are intentional from those which are not? The answer that I shall suggest is that they are actions to which a certain sense of the question "Why?" is given application; the sense is of course that in which the answer, if positive, gives a reason for acting. (Anscombe 1963, p. 9)

So, to deny application of that question to something I am doing is to deny that I was doing that intentionally. Moreover, Anscombe includes intentional actions in the class of things known without observation. According to her, knowing that I am doing something by observation leads to denying application to the question and is then a clear sign that I was not doing it intentionally: "By the knowledge that a man has of his intentional actions I mean the knowledge that one denies having if when asked e.g. 'Why are you ringing that bell?' one replies 'Good heavens! I didn't know *I* was ringing it!'" (Anscombe 1963, pp. 50–51). In this example, I am not saying why I was ringing the bell; my answer involves rather rejecting the question as lacking application to this case. I think that Anscombe's point would be stronger against possible counterexamples if she had put it in a slightly different way, perhaps by saying that we do not know what we are doing intentionally by *discovering* that we are doing it, or that this knowledge does not *surprise* us.¹²

Now, how does this apply to the problem of self-knowledge in the case of beliefs arrived at through conscious deliberation? Though, as

¹² I say this because I think it is quite likely that observation is involved in our knowledge of many things that we intentionally do. If, for instance, playing the guitar is an intentional action of mine, it seems that I am monitoring and knowing about this by observing some surfaces of the guitar, the strings, my own fingers and maybe other things in the particular context.

Moran himself remarks, Anscombe does not explain why intentional actions are known in that special way, he none the less conjectures an explanation. He writes:

The description under which an action is intentional gives the agent's primary reason in so acting, and the agent knows this description in knowing his primary reason. This description is known by him because it is the description under which he conceives of it in his practical reasoning. It is the description under which the action is seen as choiceworthy by him, as aiming at some good to be achieved. (Moran 2001, p. 126)

It seems doubtful to me that the description under which someone's action is intentional gives her primary reason to act that way, at least if we understand "primary reason" in a roughly Davidsonian way, as consisting of a pro-attitude and a belief. The description "cutting a piece of wood" can be a description under which my action is intentional, but knowing that description is not enough to know my primary reason for doing it. It is rather in knowing my primary reason (say, that I want to repair an old table and that I believe that *cutting a piece of wood* is a means to that end) that I know that description, as Moran himself goes on to say in the preceding quotation. As we have seen, that description appears in the agent's practical reasoning as the content of the instrumental belief.

How would this apply to the knowledge of avowed beliefs that are formed through explicit deliberation, as is the case in Evans' example? A plausible response might go as follows. Avowing and endorsing a belief are intentional actions themselves,¹³ which, following Anscombe, I would know without observation (or discovery). This could be correct but, as we have seen, it does not explain how this knowledge is possible. Now, applying the line of explanation that Moran suggests concerning non-observational knowledge of our intentional actions to the case of beliefs formed and avowed as a result of deliberation, knowledge of these beliefs would owe to the fact that the corresponding belief, as it is avowed, would appear in the agent's deliberation, maybe as entertained or considered in the process of weighing up reasons for and against it. Of course, given that the agent is deliberating, it will not appear as *the* belief

¹³ A related suggestion can be found in Wilson 2004. According to him, for Moran, "the forming or re-affirming of beliefs or intentions are themselves standardly basic actions" (p. 444).

that she *intends* to form or endorse; but this can also happen in the case of practical deliberation, especially when the agent is trying to choose between two or more incompatible ways of acting. The parallelism between action and belief is strongly suggested by Moran himself in the context of his discussion of Anscombe's views.¹⁴ Here is a sample text:

In belief as in intentional action, the stance of the rational agent is the stance where reasons that justify are at issue, and hence the stance from which one declares the authority of reason over one's belief and action [...]. It is an expression of the authority of reason here that he can and must answer the question of his belief or action by reflection on the reasons in favor of this belief or action. (2001, p. 127)

Now, which are the problems of this attempt to explain knowledge of the beliefs we form through deliberation on the model of our knowledge of those intentional actions we perform as a result of practical reasoning? Suppose, as we have seen Moran does, that we know the descriptions under which our actions are intentional in that they are the descriptions under which we conceive of them in our practical reasoning. Then, I would think, our knowledge of our intentional actions would depend on our knowledge of our reasons for performing them. But, insofar as reasons are typically beliefs and other psychological attitudes (together with their contents), this explanation assumes self-knowledge rather than accounting for it. This is not an objection if what we are trying to explain is knowledge of our intentional actions. But it *is* an objection if what we are attempting to account for is self-knowledge, immediate, authoritative knowledge of our avowed beliefs. The account would seem to be circular: if, according to it, we know these beliefs by virtue of knowing our reasons for forming and holding them, which, to a large extent, consist in turn of other beliefs, self-knowledge, immediate and authoritative awareness of such reasons, is taken for granted rather than explained.¹⁵ On the other hand, taking these reasons or beliefs to be known by virtue of being themselves part of a prior deliberation is to initiate a vicious regress.

¹⁴ In the same line, Lucy O'Brien writes: "Moran holds that knowledge of our actions and our beliefs are in important aspects the same. I think he is right" (O'Brien 2003, p. 382).

¹⁵ In a different form and context, Shoemaker has also raised an objection of circularity against Moran's view of self-knowledge. Cf. Shoemaker 2003, pp. 397–398.

To this objection it might be replied that, in order to deliberate about which action to perform, it is not actually required that an agent *knows* that she has certain beliefs; it may be sufficient that she actually *has* them.¹⁶ And something similar might apply to deliberation directed to formation of belief. My response would be as follows. It may well be that *for deliberation* it is enough to have certain beliefs. It may not be required that one *knows* one has them as well. The question, however, is whether this is also sufficient *for self-knowledge*, that is, for immediate, non-inferential, authoritative knowledge of the belief arrived at through the corresponding deliberation. And here the answer seems to be that it is not, and that knowledge of the corresponding reason-giving beliefs is also required for immediate and authoritative *knowledge* of the beliefs one forms through deliberation, at least if Anscombe's remarks about intentional action are to be of some help in understanding knowledge of such beliefs. Remember that, according to Moran himself, an agent knows the description under which an action of hers is intentional "in *knowing* his primary reason" (Moran 2001, p. 126; my emphasis), and not merely in having it.

A second reply to the charge of circularity is that it may be correct if reasons are beliefs (and other propositional attitudes).¹⁷ But it is not clear that reasons are beliefs, and that Moran takes them to be so. In fact there is, in recent times, a growing tendency to reject a Davidsonian view of reasons as psychological attitudes, as pairs of beliefs and desires,¹⁸ and Moran might well belong to it. According to the alternative view, reasons would not be mental states, or psychological entities of any kind. They would rather be states of affairs or aspects of reality. My reason for staying at home instead of going out for a walk is that *it is raining*, not that *I believe* that it is. A reason to act would be *what we believe*, rather than the fact *that we believe it*. Now, if reasons are not beliefs, then the charge of circularity would seem to lose its foothold, for then we do not need to explain our knowledge of one kind of beliefs, namely those formed through deliberation, in terms of our knowledge of other beliefs, namely those which function as reasons in this process. It would be enough, so the objection might proceed, if we knew of some fact or state of affairs that provided a reason for a certain belief. There

¹⁶ I owe this objection to an anonymous referee for *Crítica*.

¹⁷ This objection has also been raised by an anonymous referee for this journal.

¹⁸ For a review of this trend, see M. Alvarez, "Agents, Actions and Reasons", esp. pp. 55–57.

are, however, several problems with this train of thought. First of all, it seems that our knowledge of those beliefs we form through deliberation is no less immediate and authoritative if we proceed on the basis of facts or truths than if we rely on falsities, provided that we sincerely take them to be truths. If we *do not* take them to be truths, or probable truths, we just do not use them as reasons. But taking something to be (probably) true is just to believe it. States of affairs, then, function in our deliberation as objects or contents of beliefs. So, even if self-knowledge concerning a belief, *B*, which results from deliberation, did not require us to know that we *believe* that *P* as a reason for *B*, but only to know (or believe we know) *that P*, self-knowledge would still be assumed concerning *what we believe*, that is, concerning the *content* of our beliefs. Now, since beliefs have their contents essentially, this amount of self-knowledge would be enough for the circularity objection to remain firmly in place.

Moran's Practical Knowledge Model of self-knowledge (the label is Wilson's) faces then serious difficulties in what would appear to be the most favourable case, namely, beliefs formed and endorsed as a result of explicit deliberation. Many of our beliefs, however, do not seem to be formed in that way. As Shoemaker points out, in Evans' example someone is asking herself whether she should believe something or not, but this case should not guide the analysis of self-knowledge, because "most beliefs are not like that" (Shoemaker 2003, p. 396). In most cases, "one's awareness that one believes [something] is not the result of deliberation" (Shoemaker 2003, p. 395), simply because a great deal of our beliefs is not formed through deliberation. This, of course, is something Moran is aware of: "[T]he role I give to the deliberative stance is not meant to suggest that most of our beliefs are actually formed through explicit deliberation or reasoning" (Moran 2004, p. 458). It is plausible to think that most beliefs that play the role of reasons in our deliberations are not the result of deliberation themselves. And the same holds for many other beliefs we actually have. None the less, we seem to have immediate and authoritative knowledge of most of these beliefs, and we seem to be able to avow many of them, in an ordinary sense of this term, not in Moran's technical sense. How would Moran's proposal fare with respect to these cases? In response to Jane Heal (2004), who, as an objection to Moran's (rationalist) view, writes that "to a considerable extent we have our minds formed for us, by a variety of factors entirely beyond our control" (Heal 2004, p. 429), Moran presents

himself as holding a quite modest claim concerning the role of the deliberative stance in the formation and maintenance of beliefs:

The modest claim is that, while most of our beliefs and other attitudes either never arrive at consciousness at all, or only do so from we know not where, the fact remains that it is possible for a person to draw a conclusion, reach a finding, determine his belief about something on the basis of his assessment of the reasons supporting it. (2004, p. 458)

Moran's contention seems to be that, with respect to those beliefs that are *not* formed through explicit deliberation and that we are simply aware of, we don't know how (that "arrive at consciousness [...] from we know not where") there is still a role for deliberation and reasons to play, for they can always be subject to a possible reflection and assessment of the reasons we have for having or keeping them, so that they, or their negations, would then become conclusions of deliberation. They could then be avowed in conformity with the Transparency Condition. Now, even if this is true, it may show at most that it is possible for us to submit our beliefs to a critical evaluation; concerning our immediate and authoritative awareness of such beliefs, however, we certainly can have it whether or not we submit them to critical assessment; so, that such a criticism is possible does not explain our actual awareness of them, if such there is. The true order seems to be the opposite: if we can submit our beliefs to rational criticism it is because we previously know *which* beliefs they are. Moreover, a possible deliberation process, unlike a real one, does not determine the belief as the conclusion of reasoning; this means that our awareness of the belief cannot be accounted for in terms of the Anscombe-Moran model of the knowledge we have of our intentional actions. Remember that Moran's suggestion was that we know our intentional actions (the descriptions under which our actions are intentional) in the way Anscombe says we do because we conceive of them under those descriptions in our practical reasoning. And something analogous would apply to beliefs formed through deliberation. Now, even if problems of circularity or of vicious regress did not arise, if this model is to work as explanatory of our self-awareness of beliefs, we need *actual* processes of reasoning and *actual* reasons: merely possible reflection and reasoning cannot do the required explanatory work. Concerning beliefs that we simply have and are *not* formed through deliberation, there are no reasons or deliberations whose knowledge would allow us to know such beliefs in the immediate,

authoritative fashion that is supposed to characterize self-knowledge. In fact, as Shoemaker remarks, in a large number of cases “one would be at a loss to say what one’s *reasons* are for believing what one does” (Shoemaker 2003, p. 395). So, it seems that our immediate, authoritative awareness of beliefs that we simply have and are *not* the upshot of deliberation is left unexplained by Moran’s Practical Knowledge Model of self-knowledge. Moreover, even concerning beliefs that we acquired through deliberation in the past and still have, if Harman (1986, p. 38) and other authors are right that we tend to remember the beliefs and to forget the reasons and reasoning paths through which we arrived at them, then, even if we were to submit these beliefs to critical examination, it is incredible that this *ex post facto* assessment could explain our immediate and authoritative knowledge of those beliefs; the possibility of such an assessment presupposes instead that we already have such a knowledge.¹⁹

Perhaps, however, this is not a faithful or charitable interpretation of Moran’s intentions. He might not be implying that a possible deliberation process could produce the same effects on self-knowledge as an actual such process. On an alternative interpretation,²⁰ Moran might be contending that, in being open to such a possible deliberative assessment of my belief that *P*, I am entitled to “declare the authority of reason” (Moran 2001, p. 127) over this belief, so that my disposition to eventual reflective revision would make my self-attribution of the belief that *P* transparent to my eventual finding that *P*. One problem with this interpretation is that the rational justification of my belief and the authority of reason over it are bought on credit, so to speak. A more serious shortcoming, however, is that, even on this interpretation, a problem of circularity remains, for it certainly seems that my disposition to revise my belief that *P* presupposes that I already know, with the authority and immediacy characteristic of self-knowledge, that I have such a belief.

In my view, these problems are related to Moran’s central interest in self-knowledge as an expression of agents’ freedom, agency and responsibility for their mental lives. I think this is a very important contribution, and I share Moran’s view that persons can become authors of their beliefs, have appropriate control over them and be justifiably held responsible for them. My objection to him is only

¹⁹ It is even unlikely that the reasons we came to consider in this assessment would coincide with our initial reasons for adopting the beliefs.

²⁰ This interpretation has been suggested to me by an anonymous referee for *Crítica*.

that he interprets such authorship, control and responsibility by giving excessive pre-eminence to will-related concepts, such as decision or choice. He is certainly not alone in grounding responsibility and control on this central role given to the agents' choice. On the contrary, this is rather the norm in modern theories of responsibility. It is true that Moran insists on truth as the aim of belief, on the authority of reason over beliefs and actions and on the openness of an agent to the final balance of reasons, thus giving expression to the value of respect to the facts in the formation of beliefs. He writes, for instance, that "the very engagement with reasons has its passive aspect too, since in reasoning I must allow the force of reasons to act upon me" (Moran 2003, p. 404). It is also true that he is at pains to reject doxastic voluntarism, the doctrine that beliefs can be adopted at will, as the basis of responsibility for them. However, in my view, he follows too closely the tracks of Sartre by conferring on the agent a final power of decision or choice over her reasons, beliefs and other attitudes of hers. Moran seems to be afraid that, if an agent forms her beliefs with a passive attitude of respect and acquiescence to the facts, if she consciously keeps her power of choice at bay in this process, she thereby loses her freedom, control and responsibility for those beliefs. At several places in his work we can find Moran making use of will-related idioms in connection with beliefs. He writes, for instance, that a person "is credited with first-person authority when we take the question of what he *does* believe to be settled by his decision as to what he *is to* believe" (2001, p. 134). And commenting on Sartre, though he rejects a voluntaristic understanding of this author, he none the less interprets him approvingly as holding "that no attitude or impulse apprehended by reflective consciousness has any right to continued existence apart from one's free endorsement of it" (Moran 2001, p. 140). Moran's references to Stuart Hampshire²¹ and to Descartes also witness the central role he concedes to the will in grounding agency and responsibility for beliefs. Again, this view has some initial plausibility in the case of beliefs that are *formed* in the light of conflicting evidence, but it is much less convincing for what respects beliefs that we simply have and that form a passing or

²¹ Moran quotes Stuart Hampshire: "I make up my mind, and decide, when I formulate my beliefs [...]. The beliefs are those thoughts which I endorse as true. I do not merely find them occurring or lingering: I decide in their favour [...]. A belief is a thought from which a man cannot dissociate himself" (Hampshire 1975, pp. 97–98; quoted by Moran 2001, pp. 113–114). And Moran interprets him as holding that "the question [an agent] asks himself about his belief or desire is normally answered by a decision rather than a discovery" (Moran 2001, p. 114).

fixed part of our view of things. And even concerning the former, it would seem to be constitutive of reasons that their weight or force does not depend on my will, so that decisions or choices have not an important role to play in cases of conflicting evidence, either. In fact, making formation of beliefs in these, and other, cases to depend on the agent's decisions leads to a *loss* of control and responsibility for such beliefs.

So, though I agree with Moran that an agent can have control over, and responsibility for, his beliefs, I tend to think that this control and responsibility have little to do with decisions or choices. Forming beliefs in a controlled way, so that they can justifiably be ascribed to me as their author, depends, to a large measure, on keeping my will and choices silent and listening instead to the facts, to what is there, as a decisive guide and source of what I end up believing. And I can certainly avow beliefs formed in this way and take responsibility for them even if they are not, in any important sense, a result of my decisions or choices. Being passive in the described way when forming and keeping my beliefs is, somewhat paradoxically, a way of exercising agency and control over them, and a ground of my responsibility for them. And this holds not only with respect to our beliefs about the world, but also to our beliefs about such beliefs and other mental attitudes of us, that is, for self-knowledge. Not all control, agency and responsibility depend on the will. We are still agents when we passively obey, defer and submit ourselves to the voice of what is there in forming and maintaining our beliefs.²²

So I tend to see the role that Sartre assigns to the agent's choice in reflection and self-knowledge as clearly excessive, and conducive to arbitrariness and lack of control over one's own mental life, rather than the other way around. And I think that Moran would do well in distancing himself from such a will-centred perspective on reflection and responsibility for our minds. This might lead him to seeing the deliberative stance toward one's own attitudes as closer to the theoretical stance, as well as to viewing the latter as not necessarily at odds with agency, authorship and responsibility. As I see things, it would be good if he were to admit the possibility of a theoretical stance toward one's own mind that did not involve pathology or alienation from oneself and was compatible with avowal of the attitudes so known and with responsibility for them. This theoretical stance is, I think, the one we in fact take toward many of our reasons for believing or acting, as well as to many other attitudes of

²² I have developed these ideas in more detail in Moya 2006, chapter 5.

ours. Besides, including this form of non-objectifying, agential-cum-theoretical stance in Moran's account of self-knowledge might also allow it to overcome some of the difficulties we have pointed to in this paper. On the negative side, however, accepting this theoretical-cum-agential stance would raise the problem of explaining what such a stance would consist in and how it would be possible; and this, in turn, would open again most of the fundamental questions about the nature and possibility of self-knowledge itself. But this resistance against successive proposals of solution is anyway what experience has taught us we should expect concerning this and any other central philosophical problem.²³

BIBLIOGRAFÍA

- Alvarez, María, 2005, "Agents, Actions and Reasons", *Philosophical Books*, vol. 46, pp. 45–58.
- Anscombe, Elizabeth, 1963 [1957], *Intention*, 2nd ed., Basil Blackwell, Oxford.
- Evans, Gareth, 1982, *The Varieties of Reference*, Oxford University Press, Oxford.
- Fischer, John M. and Mark Ravizza, 1998, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, Cambridge.
- Hampshire, Stuart, 1975, *Freedom of the Individual*, Princeton University Press, Princeton.
- Harman, Gilbert, 1986, *Change in View*, The MIT Press, Cambridge, Mass.
- Heal, Jane, 2004, "Moran's *Authority and Estrangement*", *Philosophy and Phenomenological Research*, vol. 69, pp. 427–432.
- Lear, Jonathan, 2004, "Avowal and Unfreedom", *Philosophy and Phenomenological Research*, vol. 69, pp. 448–454.
- Moran, Richard, 2004, "Replies to Heal, Reginster, Wilson, and Lear", *Philosophy and Phenomenological Research*, vol. 69, pp. 455–472.
- , 2003, "Responses to O'Brien and Shoemaker", *European Journal of Philosophy*, vol. 11, pp. 402–419.

²³ Research for this paper has been funded by the Spanish Ministry of Education and Science, in the framework of the project "Belief, responsibility and action" (ref. HUM2006–04907/FISO). The first version of the paper was presented at the XVI Inter-University Seminar on Philosophy and Cognitive Science, which took place in Valencia, Spain, in March 2006, and was devoted to Richard Moran's work. I am grateful to the participants in this Seminar for their comments. I thank especially Moran himself for his comments on my contribution during the seminar and for generous additional help afterwards, though I am sure he does not agree on my interpretation of his work. I am also grateful to two anonymous referees for *Crítica* for helpful comments and criticisms.

- Moran, Richard, 2001, *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton University Press, Princeton.
- Moya, Carlos J., 2006, *Moral Responsibility: The Ways of Scepticism*, Routledge, London and New York.
- O'Brien, Lucy, 2003, "Moran on Agency and Self-Knowledge", *European Journal of Philosophy*, vol. 11, pp. 375–390.
- Perry, John, 1979, "The Problem of the Essential Indexical", *Noûs*, vol. 13, pp. 3–21.
- Reginster, Bernard, 2004, "Self-Knowledge, Responsibility, and the Third Person", *Philosophy and Phenomenological Research*, vol. 69, pp. 433–439.
- Shoemaker, Sydney, 2003, "Moran on Self-Knowledge", *European Journal of Philosophy*, vol. 11, pp. 391–401.
- Wilson, George M., 2004, "Comments on *Authority and Estrangement*", *Philosophy and Phenomenological Research*, vol. 69, pp. 440–447.

Received: April 6, 2006; revised: November 2, 2006; accepted: December 13, 2006.