The Authority of Self-Consciousness

Richard Moran
Harvard University

Since the publication of Self-Knowledge and Self-Identity in 1963 and over the course of several papers since then, Sydney Shoemaker's work has been central to virtually all contemporary thinking on self-consciousness and first-person authority. And a good measure of its importance has been not only as an evolving philosophical account of these phenomena, but also as a model of an account that places the capacity for specifically first-person awareness of one's mental states at the center of what it is to be a subject of mental states in the first place. For not every philosophical account of introspection will take its specifically first-person features to be essential to it (e.g., if it is allowed that this "faculty" could in principle be directed at the mind of another person), or seek to account for the capacity for self-knowledge in terms of the requirements of mentality and rationality generally. One can well imagine a philosophical account of how it is possible for there to be, say, non-observational awareness of one's own states of mind, awareness that does not base itself on evidence, behavioral or otherwise, and yet for that account to describe that capacity as something more or less external to the cognitive capacities involved in the having of mental states themselves. If self-knowledge is thought of as a kind of "internal scanning," then we may conceive of creatures with the full range of mental states and capacities, but who lack this extra cognitive capacity. On such a view there would be nothing essential to the having of mental states themselves which
involved the capacity for specifically first-personal awareness of them. And, indeed, the contingency of the capacity for self-knowledge (assuming a creature with mental states) is an intended consequence of naturalistic accounts of introspection, such as those of David Armstrong and D. H. Mellor, which compare it to a form of sense perception.3

Shoemaker’s own account is certainly not an anti-naturalistic one (at least on some central versions of what naturalism requires), but a consistent theme from his earliest writings on self-knowledge has been the criticism of accounts of introspection which model it as some form of (internal) sense perception.4 And again, the importance of this issue is not simply over the correct understanding of how non-observational awareness is possible, or how introspection “works,” but also over the question of whether this capacity is an essential part of what it is to be a subject of mental states, or whether it is a capacity we simply happen to have and whose absence would, in principle, leave the rest of mental life unaltered.

In this paper I want to explore a general type of argument, at work in Shoemaker’s writings and appealed to by other philosophers, against the idea that there is at most a reliable contingent connection between being a subject of propositional attitudes and having the capacity for self-consciousness of those attitudes.5 But I will be using the examination of this type of argument to suggest some revisions in our understanding of the target notions of “conscious belief” and “first-person authority.” In a sense I am trying to push further Shoemaker’s sustained criticism of the Perceptual Model of self-knowledge, and in so doing I am aware that I am probably pushing things in a direction he would prefer not to go.

I

A number of philosophers have argued from the systematic rationality of attitudes like belief to a requirement of “first-person accessibility” with respect to them. If it could be shown that being a subject of beliefs and other attitudes in general entailed having special first-person access to them, that would be an important part of the case against thinking of this kind of access as a contingent quasi-perceptual capacity for “internal scanning.” For this would mean that the capacity for self-consciousness was not a faculty that, as it happens, is “added on” to the cognitive capacities that go into being a subject of propositional attitudes, but is already an essential aspect of being such a subject.

Early in “Rationality and Self-Consciousness,” Shoemaker presents one version of the general idea.

Part of the case has been well stated by Colin McGinn. To quote from him, “If a person were not aware of his beliefs, then he could not be aware of their inconsistency; but awareness of inconsistency is (primarily) what allows normative considerations to get a purchase on beliefs; so the rational adjustment of beliefs to one another seems to involve self-consciousness, that is, knowledge of what you believe.” McGinn goes on to suggest that “the very possession of propositional attitudes requires self-consciousness: for the possession of propositional attitudes requires sensitivity to principles of rationality, and such sensitivity in turn depends on awareness of one’s attitudes”. This is, of course, precisely the view I wish to defend.6

It is certainly natural to think that the capacity we have for awareness of our own beliefs and other attitudes (“self-consciousness” in that sense) is bound up with our rationality in some way, and with the rationality of the attitudes to which we have this access. One immediate reason for this would be the apparent irrationality that often characterizes attitudes which are under repression of some kind, and to which we do not have immediate introspective access.7 And on a natural picture of the function of self-consciousness, this connection between repression and impaired rationality makes perfect sense. If some attitude of mine is out of reach of the light shone by consciousness, then I cannot “see” it in order to check it out and adjust it to the rest of my beliefs and desires. As with the intelligent, directed control of other things, the rational adjustment of one’s beliefs and other attitudes requires, at the very least, awareness of what is being subjected to control.

But at the same time, we might wonder about the naturalness of the picture of this requirement. For there is more than one thing we might understand under the heading of “adjustment” and “sensitivity to reasons.” There is no need to challenge the idea that the concept of belief (and that of other propositional attitudes) is of something systematic and normatively controlled, so that having a belief involves having a network of other beliefs, which are mutually rationally supportive. But if what McGinn means by “the rational adjustment of beliefs” is simply whatever it is that keeps one’s belief-system in good working order, avoiding gross inconsistency, updating and revising beliefs in the light of the constant flow of experience, then there is no reason this regulation could not take place below the threshold of consciousness, perhaps at a wholly “sub-personal” level. The purposes of belief-regulation do not require that the person get involved here at all, and hence there is no need for the person to have awareness of the contents being regulated. In this sense, there is no more reason to think that the rational regulation of belief requires awareness of the beliefs in question than there is to think that the regulation of heartbeat, respiration, or metabolism require conscious monitoring, or even a creature with the
capacity for consciousness at all. And as the comparison suggests, not only do such activities not seem to require the conscious control of the person, but it seems that they could not require this, for it is difficult to imagine how things could proceed if such monitoring were needed for proper functioning. In our biological as well as our cognitive lives, it must be the case that the majority of these processes, however "rationalizable," take care of themselves. Otherwise we couldn't get on with reflective activities of the whole person like walking or self-criticism. It may be true, even trivially true, that I cannot revise my beliefs unless I am aware of them. But my beliefs, like the flow of perceptions, interact and undergo revisions all the time without any intervention on my part. Their general rationality does not need my constant supervision. Indeed, it is not clear that it would even be coherent to require such global supervision as a general condition of rationality.\(^6\)

In addition, there is an ambiguity in McGinn's reference to the "control of logic over thought," which makes it harder to see that there is a question here as to why an inconsistency should be any harder to maintain once the person as such is made aware of it. In the continuation of the passage quoted by Shoemaker, he says, "once such a conflict comes to awareness one or the other belief must go — normative considerations then operate to determine your beliefs. . . . Without such self-consciousness, the control of logic over thought would be deprived of its compelling force." (20). Both sentences employ a language of necessitation here ("must go," "compelling force"), and so we might understand these claims normatively, as statements of what reason or logic requires. Logic requires the abandoning of one of two conflicting beliefs, for what is meant by "conflict" here is precisely that they cannot both be true. This, however, is a requirement of the logic of belief regardless of whether the beliefs in question are conscious, unconscious, or whatever. The normative force of the laws of logic is not dependent on the belief as an object of awareness at all. On the other hand, if we interpret the language of necessitation in this passage more purely psychologically, as a description of what must actually happen in the life of a believer, the claim is hardly true without exception (more than one interlocutor has cheerfully maintained (two conflicting beliefs after some Socrates has pointed out their inconsistency). And to the extent that it is true, it remains something of a mystery why in such a case consciousness should make such a difference, especially since in plenty of other situations of life we successfully avoid inconsistency without any awareness of the beliefs in question. The necessitation, or control of logic over belief, is often strongest when no explicit reflection is involved.

(The ringer on the telephone is turned off. I know this because I turned it off earlier. But now the phone rings, and I pick it up and say hello. I might have remained stuck in self-contradiction, but mercifully my beliefs have revised themselves and I can assume—if I bother to—that I misremembered turning off the ringer. Meanwhile, life goes on.)

Further, if what is meant by the "rational adjustment of beliefs" is not just whatever processes, personal or subpersonal, manage to maintain order, but the deliberative activity of the agent, then the idea that this requires the awareness of one's own beliefs raises a different set of questions, and these will occupy most of the rest of the paper. Briefly, I will be arguing that, whereas awareness of one's beliefs is too strong a requirement on "rational adjustment" when this is interpreted in the first way (roughly, the sense consistent with "functional regulation"), it is too weak a requirement when "rational adjustment" is understood in this second way, as a description of the person's deliberative relation to his beliefs. For the awareness of a person's beliefs, even one's own, can be arrived at in various ways, and nothing in the argument as given takes us to an account of, or an explanation of, the need for, specifically first-person awareness of one's beliefs. The picture of the need for self-awareness here does not distinguish specifically first-person awareness from the sort of awareness of an attitude that any other person would require if he wanted somehow to alter one's state of mind, a need much the same as the sense in which I need to see the furniture in my room if I am going to rearrange it. By contrast, I would like to show how we need an account of what it is for a belief to be conscious, or for one to be self-consciously aware of one's belief, which goes beyond having second-order awareness of it, even when that awareness is not mediated by behavioral evidence, or evidence of some other kind. Along with Shoemaker I think the connection with rationality is indeed crucial to understanding the first-person accessibility of one's beliefs, specifically the "immediacy" or independence of evidence of first-person reports of belief; but I think that seeing what "immediacy" comes to, and why it matters to rationality, requires some revision both of the notion of "conscious belief" and of the kind of authority involved in first-person statements of belief.

II

When we speak of "authority" in connection with first-person statements of belief and other attitudes this idea has various dimensions. There is, of course, the epistemic authority of the report, indicating that the person making the report is in a superior position to know; and hence we can see the idea of "authority" as describing the relation of the person to the facts reported, facts about a particular person's state of mind. But there is also a sense of "authority" as applied to the relationship of the person to the report
itself, indicating the person as responsible for the report, as its author. And it is not just the report that the person is author of, but also, in a central range of cases, the person can be seen as the author of the state of mind itself, in the sense of being the person responsible for it. This is the person responsible for its justification and coherence with the rest of what she believes, and whose commitment to its truth makes for the difference between this state of mind being one of belief, rather than a wish, a supposition, or a passing thought. One idea to be explored, then, is the connection between the ideas of knowledge of oneself and responsibility for oneself.

To begin to see this, we might consider situations where normal first-person authority is lost or suspended. One such situation might be the case of certain forms of psychotherapy (including psychoanalysis), where it is a working rule of the therapy that the client’s word is not taken as authoritative for the actual identity of his attitudes and states of mind. What he says about his attitude toward his mother is, as it were, simply more data, more behavior for interpretation, from which, together with other information, his actual attitude may be inferred. The therapist takes the person’s report of his state into account, but only as a piece of evidence not itself any more authoritative than various other aspects of behavior and history.

And, of course, the client himself may be well aware that his words are being taken this way. But now, if the person, being aware of this aspect of the situation, can only say what he seems to believe or feel about his mother, his limitation here is not only that he cannot see clearly into the contents of his mind and report what’s there. For we can try to imagine the abrogation of first-person authority as quite generalized, covering all questions about his current beliefs. If the client can only describe what he seems to believe about any matter, then if he is, for instance, asked some factual question that requires some figuring out, he cannot even take the answer he delivers necessarily to report the actual belief he has arrived at. What he announces confidently as the conclusion of his thinking is one thing (“Yes, I do think you save time if you avoid the tunnel, but it’s easier to get lost”), but his actual belief, as an empirical psychological matter, is another. The radical abrogation of first-person authority means that he cannot take for granted that the conclusion he arrives at is, now, what he genuinely believes about the matter. Thus it is not only that the current of his true beliefs and feelings runs somewhere out of sight of his consciousness, but also that this current seems to run its own course and have nothing to do with his explicit thinking about the people and things his feelings are directed upon.

In this way, the authority that is lacking here is not simply the epistemic authority of the truthful observer, but also the authority to make up his mind and have that actually count for something. The loss of this authority is not made good if supplemented by expert knowledge from the outside of what his real attitudes are, knowledge that might be offered to him by the therapist. Just being provided with the information about his belief is not enough. For if he cannot see the empirical question of what he believes as answerable to his current explicit thinking about the matter, then just being informed what his belief is leaves open the question whether this shall count as a reason for him or not. As it is, he is no better off than if he had been told that some other person has this belief, or that he himself did at some other time. This suggests that the 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, that it is up to him. In declaring his belief he does not express himself as an expert witness to a realm of psychological fact, so much as he expresses his rational authority over that realm.

Some philosophical accounts of first-person authority reject the idea of cognitive advantage or epistemic access altogether, and understand the authority granted to first-person statements to be purely a matter of social concessions granted to a certain class of statements. One thing we can now see wrong with the idea of first-person authority as a matter of social concessions is that it gives a misleadingly permissive picture even of the social context. It is not, after all, simply that we allow what people say about their current state of mind to go by without the benefit of evidence and are reluctant to challenge what they say. Rather, the special first-person accessibility of mental states seems not just something we grant to people, but something that is a normal rational expectation we make of them. That is, an examination of the first-person should account for why someone’s need to rely on behavioral evidence to report on his mental states would suggest something wrong with him, some state of disassociation, and would raise doubts about the rationality of those attitudes of his which are not accessible to him in the normal “immediate” way.

This sense of first-person authority as demand, rather than concession, also emerges in those situations where the other person’s feelings are crucial to your own well-being. If (out of some misguided effort at realism or sincerity) your companion will only tell you what he thinks he feels, to the best of his knowledge, about your plans together, you cannot be blamed if your response to this is to feel let down. But it’s not really epistemic disappointment that you are experiencing here, for it wouldn’t repair the situation for him to add that, given the psychological evidence, he really feels quite confident about this attribution to himself. (About some matters, you really need to feel that you’re talking to the person in charge.) The emotional involvement with the other person demands that he be in a position to speak for his feelings, here and now, and not speculate about them more or less accurately.
We begin to see the limits of the standard, purely theoretical account of first-person authority, even before we consider particular philosophical accounts of self-consciousness. For there are difficulties at the very beginning, with how the target notion is understood, the notion of a belief that is the object of self-consciousness, or, as we say, a "conscious belief." And I believe that criticizing this starting point already takes us some way toward showing the need for an alternative account along the lines I will sketch, an account which is neither deflationary nor purely a matter of good cognitive or theoretical access to one’s state of mind.

It is a commonplace in discussions of self-consciousness to conceive of the target notion in terms of second-order states. David Mellor’s paper “Conscious Belief,” for example, tries to build this assumption into an entire account of the phenomenon in question.11 Mellor begins with what he calls an “action theory of belief,” which needn’t be disputed here, since it can be understood to stand for the basic assumption that the notion of belief (and related states) is tied to its role in the explanation of behavior. This assumption is most directly opposed to the idea that belief and the like are intrinsically identifiable phenomenal states, a view which has few adherents today. Now, it will follow from such an action theory of belief that at any moment a person may be described in terms of a host of dispositional states, (e.g., assumptions taken for granted), as well as explicit beliefs and desires which together contribute to his immediate thinking and behavior. Not all of this will (or even could) be conscious at any time; some of it never will. Hence we need a term to describe “the new state of mind I come into when a belief of mine becomes a conscious one” (88), and Mellor settles on the term “assert.” What he describes as his main thesis, then, is “that asserting to a proposition is believing that one believes it” (90).12

Part of what drives the analysis of conscious belief in this direction is simply the fact that “being conscious of” is undeniably a cognitive relation of sorts. It is a way of knowing something, or having it available for further thought. But, nonetheless, I think it can be seen that, with respect to the awareness of mental phenomena, the idea of second-order beliefs is too broad to capture either the particular character of conscious awareness or the specifically first-person character of conscious belief. As to the first point, consider the ordinary phenomena of either tacit or unconscious beliefs. A tacit belief may be something which the person takes for granted but has never reflected on explicitly. Contrary to what Mellor suggests (93), such a state is not simply a disposition to assert to the proposition in question (a disposition which, since it doesn’t require the cooperation of a desire in order to assert itself, Mellor would place outside the class of genuine beliefs). Rather, tacit beliefs can be perfectly real beliefs, which interact with relevant desires to produce action in much the same way as do explicit beliefs. Problem solving (a desire-guided action) often requires making explicit some tacit assumptions which stood in the way of a solution. The person’s action in pursuing a particular misguided line of thought is explained by reference to his maintaining the faulty assumption, and explanation of his action requires ascribing this assumption to him. And, far from the tacit belief’s being a mere disposition to assert (as on Mellor’s account), it’s coming to consciousness in such cases is accompanied by immediate dissent from the proposition one had been taking for granted.13 It was thus a belief maintained only on condition of its not being a conscious one. Tacit beliefs of this sort, then, are genuine beliefs.

Now, if it’s agreed that first-order beliefs can be either unconscious or tacit beliefs, then there’s no reason why the same cannot be true of second-order beliefs too. Thus, for instance, a person may take it for granted that his friend, like most people, believes that dead men tell no tales, even if this particular thought about his friend’s admirable belief has never crossed his mind. So, in this case, he has a tacit second-order belief about his friend. Now, of course, he may equally well take it for granted that he himself does not differ from his friend in this respect, again without that thought’s ever occurring to him. In both cases, then, he has a belief about someone’s belief without its ever occurring to consciousness. And in the second case it’s not just a belief about someone’s belief, but it is also a self-directed second-order belief, which, for all that, is still not a conscious belief. Similar considerations apply to the case of beliefs which are not tacit, but which are unconscious for perhaps more psychologically interesting reasons. It would seem, then, that the particular features of conscious awareness of belief (what Mellor calls "assert") cannot be described merely in terms of second-order beliefs; not when my belief concerns someone else’s mental state, nor even when it concerns my own.

However, more important than the failure of such an account to capture the idea of explicit awareness in the account of conscious belief is the absence in it of any sense of what sort of difference is made by the distinctively first-person awareness of one’s belief. We may think of the problem in the following way. We saw that I can have a second-order belief whose object is someone else’s first-order belief, without that involving for either of us any episode of explicit awareness at all. What I now want to point out is that an analysis like Mellor’s is not rescued by refining it so as simply to require explicit, episodic awareness of belief. The particular first-person character of conscious belief would still be missing. For, of course, I could be explicitly, consciously reflecting on my friend’s belief about life on Mars without that making it a conscious belief of anyone’s. Nor need it make any
essential difference if it were my own belief that I was consciously reflecting on, if, say, I ascribe it to myself under a name or description I don't recognize. Nor even if I knowingly attribute the belief to the person I recognize as myself, using the first-person pronoun, but, say, ascribe it to myself only on the basis of reading a diary entry I wrote some time ago (where I have no reason to think I've changed my mind, even though I can't now remember why I believed this in the first place).

What any of this would leave out is the fact that to call something a conscious belief says something about the character of the belief in question. It is not simply to say that the person stands in some relation of awareness to this belief. A purely theoretical model of self-consciousness suggests a picture of awareness directed within, as much as possible like awareness directed outwardly. But compare: If someone is looking at a tree, referring to it then as an "observed tree" would not express anything about its qualities as a tree, and the same goes for the unspecified awareness of someone's belief. By itself it does not characterize the belief as such to refer to it as the object of someone's awareness, even when it is the person's awareness of his own belief. By contrast, a conscious belief enters into different relations with the rest of one's mental economy and thereby alters its character. We speak of the "consciousness" in the phrase "conscious belief" as something that informs and qualifies the belief in question, and not just as specifying a theoretical relation in which I stand to this mental state. If it were simply a special immediate theoretical relation I have to this belief, then there would be no reason in principle why another person could not bear this same relation to my belief. (That is, we could imagine for a moment that something like telepathic awareness is possible.) But in such a case my belief would not thereby acquire the attributes we have in mind when we apply the term "conscious" as a characterization of the belief itself. What I am suggesting is that such telepathic awareness would not render my belief a conscious one even were I to direct this faculty at my own belief.

I've been trying to bring out the difference between a belief which is simply the object of awareness, something one may be consciously aware of, and what we mean when we speak of something as a "conscious belief." When a belief of mine is conscious, it is not some opaque phenomenon which I am somehow aware of. I do not simply ascribe to myself. Rather, I see myself in this belief; my conscious belief forms the basis for my further train of thought about the thing in question. A somewhat more extended case may help to sharpen the sense of the inadequacy of the purely theoretical perspective on conscious belief.

It is not only what we have been calling "tacit" beliefs that do not become conscious ones simply in virtue of being the objects of one's second-order beliefs. Think of something like an attitude of resentment, which we think of as unconscious not because it is tacit, but because it has been repressed. We understand one of the intermediate goals of psychotherapy to be the lifting of the repression, with the result that the unconscious attitude becomes a conscious one. How then are we to describe the state that is aimed at? It is a commonplace of the practice of psychotherapy that this goal is not to be realized simply by telling the client that he has this attitude of resentment. What is less clearly understood is why this is not sufficient, and what this tells us about both the nature of consciousness and of repression. Suppose the client is not just told this about himself, but he then believes it without any residual doubts, and he fully understands and shares the reasons for this ascription of the attitude to him. He can then attribute the resentment to himself, and identify and anticipate the situations where it manifests itself. He thus knows all about this attitude of his. All of this is clearly compatible with the attitude's remaining under repression, with its never becoming a conscious one. Despite his knowledge of it, the development and expression of his resentment against someone follows its own path, impervious to however baseless he may find it to be, gathering strength from sources he takes to be irrelevant to the question of whether this person is in any way deserving of his resentment. His awareness of it so far is not essentially different from the theoretical knowledge he may have of another person's unconscious attitude of resentment (which, of course, would not suffice to make it a conscious one). As an attribution to himself, it is still "externally" applied, and the client need not see himself in the attitude he attributes to himself.

It is important that what's meant by "external" here is not dependent on the fact that he makes the attribution on the basis of believing the report of his analyst, or by assessing the evidence about himself. To see this, we can drop the pretense that there have to be two people involved here in the first place. For we could just as well talk about the analyst himself as the one with the unconscious attitude of resentment, but now both his theoretical expertise and his skill at applying it enable him to attribute this attitude to himself more or less immediately, without any laborious theoretical inference from the behavioral evidence. What's missing, then, for the attitude, the belief itself, to be a conscious one? The description so far may suggest that what distinguishes the case from one of genuinely conscious belief is the inferential basis of the attribution, and its foundation in some body of evidence. When a belief is conscious, it will be said, my access to it is "immediate," and not based on anything else. I don't so much want to disagree with this as ask what "immediacy" means here, and why it should be thought to matter. If the "immediacy" of conscious belief means that I do not have to infer the presence of the belief from anything else, the importance of this cannot simply be the greater convenience of beliefs of mine.
which I can be aware of without a lot of theoretical construction. After all, there are plenty of inferences we make without a second thought, hardly even a first thought. And the analyst as we are imagining him may be so expert about his own case that his ascription of this attitude to himself seems just as effortless and immediate as the everyday judgment that something before me is a chair. Nor can the importance of immediacy reside in some claim to the greater certainty of immediate judgments over those that base themselves on evidence of some kind. Judgments which are immediate in this sense are not infallible; they are just not based on anything else. And plenty of our judgments which are based on evidence are among the most certain judgments we make, more certain than some of our “immediate” judgments.

IV

More recently, David Rosenthal has been one of the most resourceful defenders of what has come to be known as the “higher-order thought” account of consciousness.15 Mellor’s theory would be a variety of such an account, for, on such a view, for a belief to be a conscious belief (and similarly for phenomenal states) is for it to be the object of a second-order state of awareness. Accounts of this general type have many adherents today, but Rosenthal has given much more detail to the view and defended it against a host of objections. But although he is mindful of the difficulties of the sort we’ve been tracking here—which we might call the problem of “merely attributional” awareness of one’s mental state—he doesn’t see them as posing a challenge to the basic motivation of a “higher-order thought” theory. He notes that “One can be conscious of being in a mental state even when we would not count that state as being a conscious state. We regard mental states as conscious states only if we are transvively conscious of them in some suitably unmediated way.”16 He illustrates the former possibility with the case of unconscious anger which I only become aware of because a friend I trust tells me what is revealed in my behavior. However, “These kinds of case require only a minor adjustment. We must specify that our transitive consciousness of our mental state relies on neither inference nor observation. Mental states are intrinsively conscious just in case we are noninferentially and nonobservationally conscious of them” (738).

Naturally, given the foregoing discussion, I don’t mean to dispute this but to ask why this adjustment isn’t ad hoc, thinking of it within the spirit and motivation of a higher-order thought account. Why should “immediacy,” defined in this way, matter at all to whether a state of mind is conscious or not? Why should it matter to the person and his relation to his mental life whether his access to it is mediated by inference or observation? I think this question should be especially pressing because, while the requirement of “immediacy” is perfectly intuitive, we must also recognize that there is little agreement about just what should count as “inferential” or “observational” relations of the sort we mean to exclude. On several contemporary accounts, “inference” is a broad enough notion to cover all sorts of unconscious, or even subpersonal, computational transitions; and both recent philosophy of science as well as philosophy of mind has made us familiar with how broad and shifting the notion of “observation” can be. Given this conceptual uncertainty, how strong or clear should our intuition be that a state of mind is a properly conscious one only when our awareness of it excludes reliance on either of these ill-understood epistemic relations?

A related context where we confront the issue of knowledge that is peculiarly “immediate” is in the agent’s knowledge of his own intentional action. A familiar point since Anscombe’s Intention is that an indefinite range of true descriptions will apply to the behavior that constitutes a person’s action, and it is only under some of these descriptions and not others that the action will count as intentional.16 Someone is pumping water from a cistern, and in the course of doing so is alsoclicking out a peculiar rhythm on the pump handle. The first is part of his aim; the second, we may assume, is not. Anscombe focuses her discussion of intentional action with the claim (which is more of a working hypothesis than a definition) that “Intentional actions are ones to which a certain sense of the question “why?” has application” (11). In the course of distinguishing the kinds of case where this question is denied application, she mentions two that are of particular interest for the present discussion. One is the familiar situation where the person simply didn’t know that the description in question (e.g., clicking that rhythm) applied to what he was doing. Perhaps that’s clear enough; if someone doesn’t know he took my umbrella instead of his own, it won’t make sense to ask him why he did this or to count it as intentional. But Anscombe adds a further restriction relevant to our question of immediacy. After introducing the idea of “non-observational knowledge” (i.e., “immediate awareness” in our sense), and illustrating it with the example of a person’s ordinary awareness of the position of his limbs, she says that the relevant question “why?” would also be refused application were the person to say, “I knew I was doing that, but only because I observed it” (14). Here we get closer to our question about “merely attributional” awareness of one’s state of mind. For while I take Anscombe’s point to be an intuitive one, we can ask the same sorts of questions about it. That is, we will all agree that a person will need to know that he is doing a particular thing if we are to ask him for his reasons in doing it, but why should the particular manner in which he acquires this knowledge matter to its status as intentional? Why isn’t the
knowledge itself enough, however it is acquired? And why should the action’s status as intentional (under this description) depend on the agent’s awareness of it having this very special basis, viz., a basis that is independent of observation, evidence, or inference?

What we might call Anscombe’s Condition on intentional action tells us that the relevant question “why” is refused application if the agent’s awareness of what he is doing (under the relevant description) is not “immediate” in our sense. Anscombe herself does not offer explicit argument for this claim, or in any case her argument is not any shorter than the entire book’s case for a distinctively “practical knowledge” as something ignored in modern philosophy’s “incorrigibly contemplative conception of knowledge” (57). But we might reconstruct the following sorts of considerations in favor of the Condition, in such a way as to relate it to the requirement that the sort of properly first-personal awareness of one’s belief that constitutes its “coming to consciousness” must be an awareness in some “suitably unmediated way.” An action is intentional under a certain description when that description (e.g., “pumping water”) can be seen as answering to some aim the person is pursuing. If the person is doing that action intentionally, then he is doing it for reasons that he takes to justify it as contributing to his aim. For him to be pumping the water for justifying reasons is for him to see that action as contributing to his aim; if he doesn’t see it as contributing to his aim, then we may conclude that he is not doing that thing for the sort of reasons (i.e., justifying reasons) that make an action intentional. 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. It is the description under which the action is seen as choice-worthy by him, as aiming as some good to be achieved. The agent takes the question of what he is doing to be answered by his decision as to what is worth pursuing, and that question is not a predictive or explanatory one to be answered by observation of himself. Hence we arrive at Anscombe’s Condition: If he can only know what he is doing by observing himself, that would be because, described in these terms (e.g., clicking out the rhythm), his action is not determined by his primary reason, is not seen by him as the pursuit of some aim. Otherwise, he would know what he is doing by reflection on his reasons, non-observationally. For an agent to conceive of himself as capable of forming an intention and implementing it (which, I take it, is necessary to conceiving himself as an agent at all), he must take his intentional action to be determined by his reasons, and thus he is in a position to know a true description of his action in knowing his reasons.

The basic point can be expressed in a loosely Kantian style, although the idea is hardly unique to Kant. The stance from which a person speaks with any special authority about his belief or action is not a stance of causal explanation but the stance of rational agency. In belief as in intentional action, the stance of the rational agent is the stance from which one declares the authority of reason over one’s belief and action. It is because of the authority of reason here that I can and must answer the question of my belief or action by reflection on the reasons in favor of this belief or action. Those reasons do not centrally concern my own desires or other states of mind; rather, they concern matters such as evidence in favor of P, or the advisability of some action: “external” matters. To do otherwise would be to take my belief or my intentional action to be up to something other than my sense of the best reasons, and if I think that, then there’s no point in my deliberating about what to do. Indeed, there is no point in calling it “deliberation” any more, if I take it to be an open question whether this activity will determine what I do or believe. To engage in deliberation in the first place is to hand over the question of one’s belief or intentional action to the authority of reason.

Of course, this authority can be partial or hedged in various ways. When I know this to be the case—for instance, when I know that I am akatic with respect to the question before me—that compromises the extent to which I can think of my behavior as intentional action, or think of my state of mind as involving a belief rather than an obsessional thought or a compulsion. Nor does a person speak with first-person authority about such conditions.

There are really two aspects to immediacy here, what we might call Epistemic and Practical immediacy. With respect to the latter, it is because the deliberator declares the authority of reason over his thought and action that at the conclusion of his thinking there is no further thing he does to make that conclusion his actual belief or his intention. He is not “working on” his states of mind, the way he might do with respect to another. He is not trying to produce results in himself. If he were, then he would indeed need to “see” what the states of mind are which he is trying to manipulate, the way the akatic person relates to his desire. Rather, he has at the outset of his deliberation delivered over the course of his thought to the authority of reason.17

We could compare this with the two ways a person might “move his arm.” In the ordinary case, he does not do any further thing in order to accomplish the moving of his arm. He simply does it, and he can do it with his eyes closed, without needing to observe anything (not on the “inside” either). In an abnormal case, on the other hand, the arm might be paralyzed and he manages to move it by picking it up with his other hand. That’s “moving his arm,” too. But here his relation to it is like the relation he has to the arm of another person, and in such cases he will need to observe what he is doing to bring it off. In the ordinary case of intentional action, I do not
“move my body” as I might move a piece of equipment, nor do I relate to my attitudes as mental items to be arranged. 18

Similarly, with respect to “epistemic immediacy,” it is because the rational agent declares the authority of reason over his thought and action that he does not apprehend himself under a particular description when he is deciding what to think or do. Several writers on self-consciousness have stressed that the target notion requires that the awareness in question should be expressible using the first-person pronoun in its “subject-use”, that is, independently of any identifying description to secure reference. 19 This is, however, part of the same requirement of “immediacy” in first-person authority. The use of the first-person pronoun in its “subject use” reflects the fact that attention is directed wholly outward, at the reasons relevant to the truth of some proposition for belief. Outside of special (and obvious) cases, it is not facts about me or what descriptions I fall under which are relevant to the truth of a proposition for belief. In deliberating about some matter I do not even take as fixed whatever stock of beliefs and desires I may bring to the problem, for entering into the spirit of rational deliberation means that I acknowledge that reflection on the problem may lead me to abandon or revise any one of them. 20

The “immediacy” of self-consciousness and first-person authority, the fact that I can be aware of my belief without inference or evidence, is a function of the fact that information about myself which I would gain through inference or evidence is ruled out as irrelevant to the question of what I am to believe. I do not need to identify a particular person since the question before me does not ultimately concern what is true for some person answering to some description, but simply what is true. (It doesn’t affect this point that of course I am dependent on my own fallible powers of reason here. That fact doesn’t alter the object of my search.) A person is credited with first-person authority when we take the question of what he does believe to be settled by his sense of what he is to believe.

V

Earlier, in criticism of the understanding of conscious belief in terms of second-order states, we were trying to imagine a kind of knowledge of one’s belief that may be immediate, but which remained spectatorial or attributional. It was argued that if the attitude in question were unconscious in the sense of repressed, coming to this sort of awareness of it, whether theoretically mediated or not, would be consistent with the attitude itself remaining under repression. There are two senses in which such an attitude may remain repressed even when the person has this kind of attributional knowl-

edge of it. First, the belief remains under repression in the sense that it fails to be a conscious belief, any more than someone’s awareness of his blood pressure turns that into a conscious process. Second, the attitude remains under repression in the sense that the person still fails to identify with it, even though he ascribes it to himself. The repressed attitude of resentment may still be alien to him in the sense of being an attitude he withholds endorsement from. In the course of therapy I may acquire the ability to identify this attitude in myself, but I do not thereby accept it as representing my view of things. To think so would be to fail to distinguish insight from acquiescence here. And because of this difference I cannot assume that the attitude of mine is sensitive to whatever reasons I may bring to bear on it. The attitude’s persistence in me is independent of my own current sense of the reasons in favor of it or against it. I cannot take for granted that considerations which I see as undermining it rationally, will actually undermine it empirically as a psychological fact about me. I can apply rational criticism to it, but only as I would to anyone else’s attitude. The resentment itself remains healthy and active, no matter how unjust, or childish, or simply unwarranted I may take it to be.

To say that the attitude remains repressed although I can successfully attribute it to myself means that I can only confront it as a kind of brute psychological datum. The kind of control I may exercise over it is something purely external, manipulating my thoughts and their expression in various ways in the attempt to redirect them. Assuming this sort of control over my attitude involves no more of an identification with it than does one’s efforts to bend or shape the attitudes of other people (where this is something different from reasoning with them).

In referring to “two senses” in which a belief or other attitude may remain under repression even when I have this kind of theoretical awareness of it, it’s clear enough that I want to think of these senses as in the closest possible relation. The difference made to the belief itself when it becomes conscious, and which is not shared by a belief that is (merely) the object of a second-order belief, is a matter of how the person relates to it as a rational agent. When I have a second-order belief about a repressed attitude of mine, though I may thereby be credited with knowledge of myself of a sort, I do not thereby see the persistence or development of this attitude as “up to me.” I may be able to exert a certain control over it through various external techniques, but I do not direct it in the ordinary way through reflection on the facts themselves (e.g., as when I alter my beliefs about someone’s capabilities simply in the course of thinking about that very person). And I am restricted to exerting this kind of “external” control over it because, although I attribute it to myself, I do not identify myself with this attitude, in the sense of seeing it as the expression of my view of the facts. Rather, I
have something like a purely theoretical, or purely psychological, relation to it.

Thus genuine self-consciousness, properly first-personal awareness, cannot be simply a matter of a special epistemic route to knowledge of a realm of psychological facts. Rather, in addition to the idea of awareness of knowledge, the (target-) notion of conscious belief itself brings with it the assumption of a particular first-personal stance toward the belief in question, the stance of rational agency. Earlier we tried to imagine a case where first-person authority was suspended across the board. In such a situation, the person's utterances and promptings could only be treated as so much data or evidence, from which his real beliefs may be inferred. On some particular occasion this could be the person's own relation to what he says or thinks about his belief or desire. We are now in a better position to see what would be incoherent in generalizing this situation.

The picture of self-knowledge usually referred to as "Cartesian" is a thoroughly perceptual one, the familiar metaphor of the "inner eye." But on one way of reading Descartes on belief and introspection, he both presents us with what I've been calling the Spectatorial picture and diagnoses what's wrong with it when he comes to explain how error is possible. That is, we may suppose there really is something like a "theater of ideas," and that each person has a kind of private access to it. But now, in the generalized situation just described, the passing show I witness really is just so much "data" to me, which it is now up to me to affirm or not as representing the facts. "By the mere intellect I do no more than perceive the ideas that are matter for judgement; and precisely so regarded, the intellect contains, properly speaking, no error." An idea as such is something I may be passive with respect to. It may be implanted in me by God or the external world or by an Evil Demon. But for that very reason it cannot be identical with my belief about some matter, for my judgments are my affair, something I am responsible for, through the exercise of my infinite liberty to affirm or deny. (God is not responsible for either the correct or incorrect use of my power of judgment.)

The problem with the idea of generalizing the theoretical stance toward mental phenomena is that a person cannot treat his mental goings-on as just so much data or evidence about his state of mind all the way down, and still be credited with a mental life (including beliefs, judgments, etc.) to treat as data in the first place. For any given mental act or utterance it may be true that I can treat it as data, something which gives me a more or less good indication of my genuine belief. But for there to be judgments or deliberation in the first place, I cannot adopt this point of view on my mental life quite generally. At some point I must cease attempting to infer from some occurrence to my belief; and instead stake myself, and relate to my mental life not as of symptomatic value, but as my current commitment to how things are out there. And so, for this reason the abrogation of first-person authority is not made up for by improved theoretical access to myself.

The agency a person exercises with respect to his beliefs and other attitudes is obviously not like that of overt basic actions like reaching for a glass. We are not agents with respect to our beliefs in the sense of being able to choose them arbitrarily. But to see the point in speaking nonetheless of "agency" here, compare the case of sensations with that of attitudes, with respect to the person's active or passive relations to them. A sensation of pain really is something one may be passively subject to, something that just happens (even though one may then bear another kind of responsibility in dealing with it). Beliefs and other attitudes, on the other hand, are stances of the person to which the demand for justification is internal. And with this comes the sense of the person's rational responsibility for his attitudes. We can fail at this responsibility in various ways, some of which I've described here, but if we lose any sense of the appropriateness of justification then we are soon no longer talking about an attitude. And in the first-person case the application of this idea is not just that of the rational appraisal of one's beliefs, but includes the responsibility for that appraisal's making a actual difference to what one believes.

Acknowledging this sort of agency and responsibility does not involve us in any sort of voluntarism about the formation of beliefs. In fact, the kind of responsibility I am trying to focus on as an essential aspect of conscious belief is strictly incompatible with the manipulation of one's attitudes as so much mental furniture. When I manipulate my attitudes through various external means it is because I do not identify with them, that is, I do not take them to be expressive of my view of things. It is because I do not in this case take them to be responsive to my best reasons that I seek to exert this sort of external control over them. In the normal case the "rational agency" I exert expresses my sense that I am answerable for my attitudes, not because I "put" them there, but because their justification is up to me, and their empirical existence as psychological phenomena is dependent on this justification. The person's rational agency with respect to his attitudes does not mean that he picks and chooses them. That would already be to express an alienated relation to his mental life. Rather, it means, for instance, that the conclusion of my practical reasoning shall count for me as telling me what I will in fact do, or that the conclusion of my theoretical reasoning shall count for me as what I do in fact now believe. This is in contrast to the situation of manipulating my own attitudes, where once I settle on what attitude to adopt I would then need to find a way to make this happen, finding the psychological means to implement my choice. Whereas, in taking the stance of rational agency toward one's thought, there is no further the
person does, after he has come to some conclusion, to make that belief or decision his own.

Sometimes, as in various conditions of psychic impairment, the claim to this kind of responsibility is hollow. And in such situations, rather than taking my declaration to settle the question (of what I really believe), I may be obliged to treat it for the moment as just so much data, something which may be a more or less good indication of my actual belief. How one emerges from such a condition is a question for another occasion, but there would be nothing that counted as agency or deliberation at all if a person could not claim the conclusion of his reasoning as making it the case that, as a matter of psychological fact, this is his belief about the matter.

The account sketched out here is incomplete in many respects, but I think these considerations provide a defense of sorts of the case Shoemaker has presented over the course of many papers against the Perceptual Model of self-knowledge. In particular, the story here proceeds by trying to get clearer about two of the most general features of Shoemaker's criticism of that model, and the kind of positive account he seeks to replace it with. First, on his account the capacity for non-observational awareness of one's own states of mind is to be understood as internally related to what it is to be a mental subject of the sort that we are (the sort that involves the concepts of personhood and rational agency). On this view it is "constitutive or definitive of mental states, or minds, or of the concepts of these, that these states intimate their existence to their possessors in a special and direct way," the denial of this constitutive relation being central to the defense of the Perceptual Model. Second, it is claimed that this constitutive relation is to be understood by reference to the involvement of self-consciousness in our status as rational agents. I've tried to bring these two ideas into some what closer relation to each other, but the defense is nonetheless an oblique one, and proceeds by way of arguing that we need to revise our conceptions of the target notions of conscious belief and first-person authority.

NOTES

1. Ancestors of this paper were delivered at the American Philosophical Association, Central Division, Chicago, 1996; Ohio State University; the University of California, at San Diego, and at Berkeley, and at UCLA. I am grateful to audiences on these occasions. Special thanks to Jennifer Church, Marcia Cavell, Luca Ferrero, Jennifer Whiting, and Bernard Kobes.


4. For some general remarks on Shoemaker's thinking concerning the differences and relations between causal and conceptual accounts, see the introduction to his 1984 collection, Identity, Causation, and Mind (Cambridge: Cambridge University Press, 1984). Within that collection, the essays "Conceptual Connections and Other Minds," "Embodiment and Behavior," and "Some Varieties of Functionalism" are especially relevant.

5. The denial of the contingency of the relation between mental states and the capacity for first-person awareness of them is an important part of the case against construing self-knowledge as a kind of perception, for perception is characterized as a contingent, causal relation between an object and a perceiver.


7. The quotations from McGinn are from The Character of Mind (Oxford: Oxford University Press, 1982), 20–21. Shoemaker's case for the connection between rationality and self-consciousness is not restricted to the line of thought described here, and in this paper I will only be concerned with the basic argument shared by him, McGinn and others. For further development of this argument, as well as independent arguments toward the same conclusion, see Shoemaker, "On Knowing One's Own Mind" (1988) and "First-Person Access," (1990), both reprinted in his collection The First-Person Perspective and Other Essays (Cambridge: Cambridge University Press, 1996).

8. Apart from the difficulties in imagining such a psychological condition, such a requirement of explicit awareness and intervention would appear to involve an infinite iteration of levels related to those the Tortoise imposed on Achilles.


12. His secondary thesis is that linguistic action—speech and writing—requires second-order beliefs.


17. Cf. Stuart Hampshire, Freedom of the Individual (Princeton: Princeton University Press, 1975): "He who employs some method or technique to get rid of an idea... acts upon himself... exactly as he might bring about an effect in the mind of another. The man who changes his mind, in response to evidence for the truth of a proposition, does not act upon himself, nor does he bring about an effect" (100).
20. A related point is made in G. A. Cohen, “Beliefs and Roles,” reprinted in *The Philosophy of Mind*, ed. J. Glover (Oxford: Oxford University Press, 1971), which defends the thesis that “to cite one’s role is never to give a good reason for a belief one holds, and that his occupancy of a certain role can never function as a belief-grounding reason in the thinking of a sane man” (54).