

Introspection and Avowable Self-Knowledge

Pacific APA, Portland 2006

Dorit Bar-On, UNC Chapel Hill

1

1. Introduction: Rorty's Dilemma

Three and a half decades ago, in a wonderful paper entitled “Incorrigibility as the Mark of the Mental”, Richard Rorty argued that a properly *realist, non-eliminativist* view of mentality must provide an acceptable demarcation of the mental from the physical. Like Descartes, Rorty thought that this required explaining the special security of subjects' ascriptions of mental states to themselves. And, like Descartes, Rorty thought that, if subjects' mental self-ascriptions are to enjoy the right kind of security, this must be because subjects are *incorrigible* about their present mental states. However, if, as Rorty himself suspected, there could be no states answering to this description, then the existence of distinctively mental states will be impugned; hence the possibility of eliminativism, a position which Rorty himself later embraced. If Rorty is right, then we may be caught between the Scylla of Cartesian Dualism and the Charybdis of eliminativism about the mental. Either we recognize the existence of mental states over and above physical states (and of Egos over and above human bodies) in our ontology, or we must give up altogether on the idea of a systematic conceptual, even if not ontological, delineation of the category of the mental.

It would be nice to be able to slip between the horns of Rorty's dilemma (as I shall refer to it). And it may seem that *contemporary materialist introspectionism* promises to do just that. On this view, we possess a distinctive mode of internal access to our own present mental states – a brain-based monitoring and scanning mechanism that allows us to obtain directly, or non-inferentially, relatively accurate information about the presence and character of these states. Although potentially fallible, corrigible, and dubitable, this mode of internal access is sufficiently reliable *and* different from other modes of access we possess to shoulder the main epistemic burden assigned by Descartes to his privileged access without invoking specially *accessible* states of immaterial substances. I want to question the success of introspectionism in this non-Cartesian incarnation as an attempt to avoid Rorty's dilemma.

Let me be clear from the start: I don't wish to deny that we possess special mechanisms that enable us to direct our attention inwardly and determine in a non-evidential, non-inferential way whether we feel a headache, how intense it is, whether we'd like some tea, what we are thinking right now, etc. It is, of course, an empirical question whether we possess such mechanisms, how reliable they are, and how sharply they contrast with the perceptual mechanisms that allow us to obtain information about things external to us – various features of our environment, including nonmental and mental states of other people. I am not concerned here to address these empirical issues. Rather, I want to ask whether introspection so understood can help us understand the special status we ordinarily assign to what people say or think about their present mental states. More specifically, I want to ask whether it can help us understand the contrast between what we may call “avowable self-knowledge”: the kind of seemingly privileged basic self-knowledge that we are often credited with when we avow, for example, “I feel dizzy”, or “I’m scared of that dog”, or “I am wondering what time it is”, on the one hand, and the knowledge we have of some of our present nonmental bodily states. For, we ordinarily presume that people’s pronouncements about their present mental states are not only more secure than what they say about *others’* mental states, but we also presume them to be more secure (even if not absolutely infallible, incorrigible or indubitable) than what they say about their own present *nonmental* bodily states. It is in good part this aspect of the commonsense view – the mental/nonmental asymmetries – that Rorty suspects to be subject to elimination once we abandon the Cartesian view.

In what follows, I will first explain the difficulties I see with materialist introspectionism *as a way out* of Rorty’s dilemma. I will then briefly sketch what I take to be a promising alternative and try to address an objection to this alternative that would seem especially pressing for anyone attracted to materialist introspectionism. What I hope to accomplish is twofold: to correct a misconception about what introspection *can* do for us, philosophically speaking, and to suggest that materialist introspectionism is not, as some believe, ‘the only game in town’ for anyone who rejects dualism.

2. Materialist Introspectionism

Consider the following *avowals*: “I’m feeling dizzy”; “My leg hurts”; “I’d like some water”; “I hope food will be served soon”. Compare and contrast these avowals with the following nonmental self-reports: “I’m spinning around”; “My legs are crossed”; “I’m sipping water”. Like avowals, these bodily self-reports, when produced in the normal way, are not based on any inference, evidence, or ordinary observation. Yet, unlike avowals, these nonmental self-reports are completely and straightforwardly open to denial or correction by potential observers on the basis of *their* observation. Looking at me, you’d be in a perfectly good position to deny and correct my self-report (“No, your legs are *not* crossed,” or “You’re not spinning around, you’re actually sitting down!”).

Authors generally unsympathetic to Cartesian Dualism are nevertheless impressed by the epistemic contrast between avowals and non-evidential bodily self-reports. Yet it is not easy to come by a satisfactory characterization of the contrast, let alone an explanation of it. As mentioned earlier, Richard Rorty has proposed *incorrigibility* as marking the contrast. The avowals of a sincere, linguistically competent subject cannot be overridden, Rorty suggests, because “[w]e have no criteria for setting [them] aside as mistaken, . . . whereas we do have criteria for setting aside all reports about everything else” (1975: 413). But this seems to me too strong, since we *are* sometimes prepared to challenge an avowal. For instance, you say, apparently sincerely, that you are not mad at me, and I wonder whether you really are not, since you seem to me pretty sulky and unfriendly. Or you say, apparently sincerely, that you’re thinking hard about the objection I’ve just raised to your argument, and I question it, because you seem very busy doing something else, and so on. In these cases, your behavior appears to give the lie to what you say, and I will not simply take your word without questioning.

Crispin Wright has suggested that what marks avowals is the fact that there is no such thing as showing oneself “chronically unreliable” in one’s avowals, whereas one’s proprioceptive ascriptions, for example, can presumably be subject to systematic global failure (1998: 17). But this claim seems too weak. For, it seems to me that the distinctive security of avowals, even if not absolute, attaches to each avowal considered individually, and not merely courtesy of a global assumption that we make about subjects. In a little while we will turn to my preferred characterization and explanation of the distinctive security of avowals, which diverges from both Rorty’s and Wright’s. But

right now I want to examine a recently popular strategy for explaining the security of avowals in a non-Cartesian way. I want to argue that in their attempt to escape the Cartesian horn of Rorty's dilemma, proponents of this strategy may risk being impaled on its eliminativist horn.

If we are no longer wedded to absolute incorrigibility as the mark of avowals, yet agree that avowals do indeed enjoy a high degree of epistemic security, it may be tempting to play down the contrast between avowals and other self-ascriptions. We can begin by noting that there is a class of bodily predicates, such as "crossed legs," or "is sitting down," whose application exhibits an epistemic first-person/third-person asymmetry. Such predicates are applied on the basis of external observation in the third-person case, but on the basis of internal perception (or monitoring, or sensory feedback) in the first-person case. Likewise, we can think of "in pain," or "is angry at y," and so on, as predicates that are applied on different epistemic bases in the third- and first- person cases. They apply on the basis of observation of behavior, or inference, or conjecture, in the third-person case, but, in the first-person case, they apply on the basis of *introspection*.

Introspection, however, need not to be understood in the Cartesian way, as an infallible faculty that reveals to each subject her own 'private' states, which no one else can access. On the contrary, *the materialist introspectionist* suggests that introspection is a faculty that reveals to us goings-on *inside* our bodies (more specifically, in our brains). On one story, it is speculated that the human brain is equipped with a special mechanism – a kind of scanner – designed to deliver reliable higher-order judgments about our first-order mental states. My distinctive ability to tell what I am thinking right now, for instance, is due to my brain's ability to scan its own present operations so as to yield highly reliable, non-evidential judgments, which are then articulated (in speech or in thought) through self-ascriptions of mental states. [REF TO BILL There are two different versions of this view. One version, known as "HOP" (for "Higher Order Perception") our scanning mechanisms deliver *perceptions* of first-order mental states. On the other, known as "HOT" (for "Higher Order Thought") they deliver *thoughts* about them. My talk of higher-order *judgments* below is intended to be neutral as between these two versions and my critical remarks are intended to apply to both.] On both these

versions, introspective awareness consists in our passing a higher-order judgment (perceptual or conceptual) on some internal going-on. Co-opting the ‘higher-order’ view for present concerns, the materialist introspectionist might suggest that avowals can be seen as descriptive reports of these internal goings-on. Avowals would then be taken to represent the higher-order deliverances of the subject’s inner detection mechanisms, and the security of avowals will be taken to derive from the epistemic reliability of these deliverances.

The materialist introspectionist view – MI for short – begins with a rejection of the Cartesian idea that there is something over and above our physical states for our avowals to report, though it tries to accommodate a suitably tempered notion of privileged access in the form of reliable mechanisms that we can deploy to obtain information about *some* of our physical states. Note right away that this strategy can at best deliver to us access that is *as privileged as* is reliable. And it can at most fund an asymmetry between avowals and nonmental bodily self-ascriptions *to the extent that* our introspective mechanisms are in fact more reliable than all mechanisms delivering information about our nonmental bodily states. The question is whether, given these limitations, MI can adequately characterize and explain the systematic contrast we ordinarily draw between avowals and nonmental self-ascriptions. In addressing this question, I will highlight various – familiar, I hope – aspects of our ordinary treatment of avowals that I think call out for a sort of explanation that MI is not fit to provide.

As I said before, I don’t think commonsense regards avowals as incorrigible. But consider a situation in which we may be convinced, on whatever grounds, that someone’s avowal is false. We still wouldn’t simply dismiss her self-ascription, as we might when denying a bodily self-ascription on the basis of our contrary judgment (“No, you are not bleeding” or “No, you’re not sitting down, you’re kneeling!”). We give considerable weight to the *very fact* that the person avowed being in the condition. We treat the person’s avowal as lying on the side of our evidence, and evidence that carries special, constraining weight. That this is so can be seen if we consider that if someone kept insisting that, e.g., she is hungry, or tired, or likes the painting, then as long as we didn’t question her sincerity, we would eventually defer to her.

Note that, according to MI, it should be perfectly possible for me to issue a false avowal due to a *brute local error* – an error that is not due to any defect of any of my cognitive mechanisms but is simply due to the fact that my scanner has been ‘fooled’ into false detection. Moreover, it should also seem entirely possible for a subject to suffer *global systematic failure*. Just as someone could lose (or may never have had) the ability to see or hear, or the faculty of proprioception or kinesthesia, so someone could lose (or may never have had) the ability to tell by introspection what mental state she is in, so that all her avowals are wildly off the mark.

On the commonsense view, though, both the possibility of brute local error and the possibility of global systematic failure seem more problematic than MI would have us expect. As regards the first possibility, note that, even when we are prepared to question the truth of an avowal, there is an assumption that the avowal’s falsity is due to some psychological irregularity, failure or defect on the part of the avower. We don’t readily accept that the subject might have simply been *fooled* into issuing a false avowal by an ‘uncooperative’ mental world. As regards the possibility of global systematic error, I’ll just note that attempts to establish systematic unreliability in subjects’ avowals through psychological experiments or in the context of therapy typically fall back on some avowals that are taken at face value and whose security is not questioned. We may question some of someone’s self-ascribed sentiments or propositional attitudes, but we do that on the strength of her *other* self-ascribed mental states. It is one thing to show that various mental self-ascriptions we make are false; it is another to show that we can make sense of the idea of a subject who, though she *has* mental states, is systematically wrong in her avowals.

Let us now go back to a sub-class of bodily self-ascriptions mentioned earlier: proprioceptive and kinesthetic self-reports. Consider again my ordinary pronouncement that my legs are crossed. Here too I rely on no evidence or inference or observation, so there is an epistemic third-person/first-person contrast in terms of the *basis* on which the pronouncement is made. But this 1st-person/3rd-person contrast in epistemic route does little to explain the mental/physical contrast, for the latter contrast is not only a contrast in the manner arrived at the self-ascription, but also a contrast in its degree of security. In the realm of proprioceptive or kinesthetic reports, my word is arrived at *differently* from

that of my observers' but it is *not clearly better* than theirs. On a given occasion, my body could clearly be in a state that my proprioceptive mechanism cannot distinguish from the state of my legs being crossed. I could also prove to be *systematically* unreliable in reporting my limb positions or bodily movements without looking. The ability to tell my bodily conditions through proprioception or kinesthesia thus seems *alienable*. Furthermore, we can even conceive of someone whose brain was so hooked up to my limbs as to receive direct information about their position. Such a person could inherit my proprioceptive ability, and would be able to tell where my limbs are the same way I do. Proprioception (as well as kinesthesia) thus seems also entirely *transferable*.

If the distinctive security of avowals were entirely due to the epistemic security of introspective mechanisms, we might expect so-called 'first-person authority' to be equally alienable and transferable. Yet I submit that by commonsense lights, it isn't. As pointed out above, the possibility of a subject with thoughts, sensations, emotions, etc. who *regularly* mis-avows seems *at the very least* highly problematic. Even more problematic is the notion that I might be able to transfer my 1st-person authority to someone else – someone who was able to ascribe to me reliably and correctly present mental states without relying on any observation or evidence. For one thing, my avowals would surely constitute a crucial part of the data against which the other's claim to be able to provide reliable non-evidential reports of my occurrent mental states would be measured. But more importantly, so long as there was no question about my grasp of the language or my sincerity, if disagreement broke out between us over what is now going on in my mind, her consistent past success in reading my mind would not be sufficient ground for taking her word over mine. It seems, rather, that we would take the disagreement as signaling the waning of her mind-reading powers. (Similar remarks would apply to Rorty's imaginary brain-scanner, which could be set up to replace one's own self-scanning mechanism and to serve as a mechanical mind-reader.)

So far, I have tried to highlight various possibilities that are allowed by MI but which commonsense intuitions take to be problematic. I have *not* aimed to offer a direct argument against these possibilities. My point has rather been that MI by itself isn't apt to explain these commonsense intuitions. Not that these intuitions are sacrosanct. But, other things being equal, I think that, in the interest of avoiding the eliminativist horn of

Rorty's dilemma, we should prefer an account that allows us to preserve them, since these intuitions underwrite not just the ordinary view of avowals' security but the very separation between the mental and the physical.

3. The Limits of Introspection

Before turning to what I take to be a promising way out of Rorty's dilemma, I'd like to offer a diagnosis for the difficulties I see with MI. The diagnosis is intended to apply to a family of views on avowals' security, which I describe elsewhere as the *epistemic approach to avowals' security*. On the epistemic approach, the distinctive security of avowals is to be explained by appeal to the contingent epistemic security of the basis on which they are made, where different epistemic views will identify different epistemic routes or methods or bases that we rely on when producing avowals. My claim is that, regardless of the epistemic route or method or basis, such explanations falter on the question why avowals should be regarded as enjoying a different security from that enjoyed by nonmental bodily self-reports. But my focus here will be on MI.

So the question in front of us is this:

How well can MI explain (or explain away) the commonsense intuitions regarding avowals' special security as it is contrasted with the security of e.g. proprioceptive self-reports (where these intuitions, as I pointed out, take avowals to exhibit not only a contrast in the manner arrived at but also a contrast in the kind of security)?

According to MI, avowals report physical conditions *inside* our bodies, whereas non-evidential bodily self-reports concern bodily surfaces, orientation, posture, limb position, etc. As a first pass, MI may suggest that introspection is simply not vulnerable to some of the failings of own-bodily perception, on the one hand, and that, on the other hand, as observers we recognize (perhaps tacitly) our relative limitations in accessing subjects' inner states. But it is hard to see why we should suppose this to help explain the kind of contrastive security we normally assign to avowals. Suppose that, as a matter of empirical fact, our brains are much better at detecting states inside our bodies than conditions at our bodies' surfaces. Are we to suppose that this empirical difference in brain mechanisms is what we respond to when assigning the kind of weight we assign to

avowals (as contrasted with proprioceptive reports)? And, even assuming that we could be credited with a tacit grasp of the relative reliability of various brain mechanisms, the higher reliability of introspective mechanisms could at best explain why we might think of avowals as *more reliable* than proprioceptive reports. It would not help explain why the authority we enjoy in connection with our avowals should be regarded as inalienable and non-transferable.

MI may point out, as a second pass, that avowals, *unlike* proprioceptive reports, concern goings-on that are normally (even if contingently) *hidden* from plain view, and are not readily observable by others. This may explain why, as observers, we regard subjects as much better placed to pronounce on their mental states than on, say, the state of their limbs. But it isn't clear how this can help with our question. After all, we do *not* take subjects to be in a privileged position to pronounce on *nonmental* conditions inside their bodies (e.g., the condition of their livers or their hearts, and many conditions of their brains). On the contrary; if someone pronounced on some hidden chemical process taking place in their spleen, we would be very doubtful. Again, let us suppose that, as a matter of empirical fact, our brains are, for some reason, more reliable at detecting a certain particular subset of our internal conditions. Still, even setting aside the question how this subset is to be delineated, the question remains why our pronouncements about only some of the conditions hidden inside our bodies should be taken to enjoy unparalleled security.

There is a reason why MI *seems* like a promising option when trying to understand avowals' security and 1st-person authority. *Assuming*, with commonsense, that avowals enjoy a special kind of security, it is natural to wonder:

(a) How *could* avowals be more secure than other pronouncements?

Consistently with materialism. And MI can be seen as offering us an empirically cogent account of how it might be *possible* for us to produce highly reliable reports about some of our internal states. But the question I have posed is rather:

(b) *Why* are avowals ordinarily treated as enjoying a different kind of security from all other pronouncements, including non-evidential bodily self-reports?

With Rorty, I take it that answering this question is mandatory for understanding the commonsense divide between the mental and the physical, and, relatedly, for developing a non-eliminativist view of mind that wasn't guilty of Cartesian ontological excesses. I submit, however, that this question is *not* adequately answered by postulating the existence of a reliable mechanism designed to deliver largely true self-reports.

In effect, question (b) demands an explanation of so-called '1st-person authority'; specifically, the fact that it is inalienable and non-transferable. MI (and the Epistemic Approach more generally) can be seen as construing first-person authority as an *epistemic* authority. Perhaps the most familiar model for epistemic authority of this sort is the model of expertise. Experts are people who 'know best.' They are individuals whom we take to have greater knowledge than most of us about certain matters, through having greater experience or training, special methods or natural facility. When we take a trained botanist to be an expert about plants, for example, we take it that she is better placed than most of us to identify or analyze certain plants, for example. So whenever she pronounces on such matters, we presume that she is right, because we recognize that she is epistemically well placed to convey more reliable information than the rest of us in the domain of botany. Note that, in general, we think of experts as people who have established themselves as knowing certain facts that *any of us could equally come to know*, at least in principle. Even oracles and prophets are presumably taken to have a special ability to know in advance things that ordinary folks *will* (or at least *could*) in due course come to know equally well, though not through *foresight*. The idea of assigning expertise to someone concerning matters that we suppose ourselves in principle unable to ascertain seems odd, even if not incoherent. So the expertise model would seem unavailable to anyone who believed in the logical privacy of mental states.

Now, although MI of course does not maintain the privacy of mental states, I think it still cannot avail itself of the expertise model. To begin with, we are ordinarily prepared to assign first-person authority in utter disregard to individual credentials; we don't wait for individuals to *establish* themselves as experts on their own mind. Also, 1st-person authority is not a matter of a special advantage that avowing subjects have regarding a *specific subject matter*. Rather, it concerns their pronouncements about what goes on in their *own* minds *at the time of avowing*. Furthermore, as suggested earlier, it is

hard to imagine how one could attempt to substantiate the reliability *or* unreliability of particular (linguistically competent) subjects concerning their present states of mind without already taking at face value many of their avowals.

On the materialist's own view, the conditions that constitute mental states, though not essentially private, are still, as a matter of fact, hidden from observers, since we do not as yet possess the necessary means for perceiving them. This means that as observers we are not, as a matter of fact, able to establish the likelihood of the truth of subjects' avowals (and thus their reliability). But this in turn means that we are *not* currently in a position to declare subjects experts on their own mental states. Yet we *do* regularly assign to subjects first-person authority. I think that all this tells against construing the ordinary assumption of first-person authority on the model of expertise. If we assign special authority to avowing subjects, it can't simply be because we make the empirical assumption that they are experts on their own present states of mind.

It seems that in order to apply the expertise model to the case at hand, we'd have to think of the mental states of others as already open to view by us. Ironically, one kind of view of mental states that fits this bill has traditionally been thought to involve an outright rejection of first-person authority and the special security of avowals. I have in mind of course behaviorist views, according to which the presence of a mental state of a certain type is in some way guaranteed by the presence of certain behavioral manifestations. On such views, it might be thought that we are, already at present, in a position to recognize that people are very good at telling what mental states they are in, and thus to declare them experts. By the same token, however, such views would require rejecting, rather than explaining, the commonsense idea of 1st-person authority as inalienable and non-transferable. (Indeed, behaviorist views are often cited as paradigms of elimination.)

4. Avowals' Security: A Neo-Expressivist Account

It's all well and good to bash MI for not giving us something we'd like. But what if MI is the best game in town? As you may guess, I don't think it is, and I have myself tried to develop an alternative, non-Cartesian, non-introspectionist account of avowals' security and basic self-knowledge. I don't have the time to offer more than the briefest

sketch of the key idea of this alternative account. In the remainder of this talk I'll do that and then address one worry about this alternative that is likely to loom large in the mind of anyone initially attracted to materialist introspectionism.

Let's observe first that proprioceptive and kinesthetic reports, *like* avowals, are "identification-free": epistemically speaking, they do not rest on an identification of the subject of the utterance or thought. That is to say, in normal circumstances, if I say (or think): "My legs are crossed" or "I'm spinning around", my utterance/thought does not rest on my separate identification of some individual as myself. I have no more reason for thinking that *someone's* legs are crossed than whatever reason I have for thinking that *my* legs are crossed. This is what renders bodily self-reports of this kind "immune to error through misidentification" (to use Evans' and Shoemaker's terminology). In the case of avowals, though, I seem to enjoy *additional, ascriptive* security: not only are avowals immune to error of misidentification, but unlike all nonmental self-reports, they are also immune to error through misascription. When avowing "I'd like some water" for example, not only do I have no reason for thinking that *someone* wants some water, other than whatever reason I have for thinking it's *I* who wants water. I also have no reasons for thinking that I'm having *some* attitude or other toward *some* substance or other over and above (or separately from) whatever reason I have for ascribing wanting water to myself.

My alternative non-epistemic account of avowals' security begins with the suggestion that, on the ordinary conception, avowals are not only immune to error through misidentification but also *immune to error through misascription*. To say this is *not* to say that avowals are absolutely incorrigible in Rorty's sense. It's just to say that avowals are immune to a certain array of epistemic errors (and thus corrections) a *much wider* array than all other ascriptions, including, specifically, proprioceptive and kinesthetic self-reports. Immunity to error through misascription, like immunity to error through misidentification, is a matter of protection from *certain kinds* of epistemic error. (Specifically, though avowals can be mistaken, when they are, this is not because they involve some recognitional mis-taking.) Whereas all other ascriptions of states to individuals, including ascriptions of nonmental states to ourselves through proprioception, are *not* immune to error through misascription. If I say or think: "My legs are

crossed” (in the normal way), then even if I cannot be misidentifying who it is whose legs are crossed, if my legs are *not* crossed, my self-report will be mistaken precisely because I have misidentified the state of my legs: I will have mistaken one state of my limbs for another. In the case of avowals, though, I suggest, I am protected from epistemic error in the ascriptive part as well. But this is not to say that avowals cannot be false, or even that we can ever be in a position to correct someone else’s avowal. In certain circumstances we may even be led to question an avowal of pain (E.G.). It’s just that when a person is avowing pain she has no reason for thinking she is in some mental state or other other than whatever reason she has for thinking that she is in pain. If her avowal is false, it will not be false due to her mistaking some other state of hers for pain due to the way the state *appears* to her. On the contrary, in opposition to MI, I maintain that the epistemic position of one who avows is *not* the position of someone who receives direct information about her present internal states, correctly or incorrectly identifying the presence and character of those states on the basis of the way they appear to her ‘from the inside’, and issuing reports of those states. Instead it’s the position of someone who is in some state and directly *speaking from* her present state.

Avowals, on my view, constitute a certain class of *acts* in which a subject gives linguistic vent to present mental states. These are acts of *speaking* her mind, instead of giving non-linguistic expression to it. Such acts are epistemically unmediated, even though they issue in true or false products (linguistic utterances, in the case of speech, articulate thoughts, in speechless cases). So in contradistinction to the materialist introspectionist view, on my account, 1st-person authority is not the epistemic authority of an expert who is able to take a super-close look at her mental states, as it were, but rather a matter of the privilege of someone who expresses the very same states that she ascribes to herself in the act of avowing.

The account I offer is designed to answer the following question:

Why is it that avowals, understood as true or false ascriptions of contingent states to an individual, are so rarely questioned or corrected, are generally so resistant to ordinary epistemic assessments, and are so strongly presumed to be true?

The question invites a search for a distinctive feature(s) of avowals that would explain why they enjoy the level and kind of security that we ordinarily assign to them. In answering the question, I try to show how it could be *reasonable* for us to treat avowals as protected from certain *kinds* of epistemic mistakes and criticisms, even if not absolutely infallible and incorrigible. And I claim that my account is able to explain better than others how avowals are different in this regard from all other empirical ascriptions, *including* non-evidential bodily self-reports. I thus claim to be able to capture and explain the ordinary notion of 1st-person authority. On my story, it is reasonable for us to assign inalienable and non-transferable 1st-person authority to subjects because this authority is not assigned on the basis of our tacit recognition of their expertise in regards to their own mental states. Rather, we assign it on the strength of recognizing their avowals as acts in which they express these very states at the same time as they self-ascribe them.

5. *Avowable Self-Knowledge*

No doubt the little that I've said here invites more questions than I've answered. But in conclusion I want to try to answer just one question that may seem particularly pressing for anyone drawn to the introspectionist view. The question is this:

Avowals aside, what explains our privileged self-knowledge?

Even forgetting about whether or not we bother to pronounce on a present mental state in speech or thought, don't we have immediate and secure knowledge of our present mental states of mind? If so, we want to find out what allows us to have such privileged knowledge. And it may seem that any account which, like my Neo-Expressivist account, focuses on the security of avowals, would seem to be besides the point. Putting it differently, it may seem that commonsense tells us not only that typically, if a subject avows (in speech or in thought) being in a mental state M, then she *is* in M, but *also* that a subject who is in M will have privileged knowledge that she is in M, whether or not she bothers to pronounce on the matter (to others or to herself). Call this the idea of 'unarticulated basic self-knowledge'.

There are different ways of understanding this idea. First, it may be read as the thought that a creature, whether human or brute, who is feeling pain or pleasure, hunger,

fatigue, anger, fear, etc., regardless of whether she says, or thinks to herself that she is in the relevant state, *knows* it, and knows it the way no one else does. This is just part of what it is to *be* in a state of this kind. If so, then, contra my Neo-Expressivist view, privileged self-knowledge is something a subject has *whenever* she is in a conscious mental state, and such knowledge is only incidentally articulated by avowals. But note that on this line of thought, “S has privileged (‘first-person’) knowledge that she is in M” is simply taken to be entailed by “S is in a ‘conscious’ state M.” And note, too, that it will turn out that any creature who has conscious mental states will *ipso facto* possess privileged basic self-knowledge. Although I think it is true that only so-called conscious states are avowable, I take the question about privileged self-knowledge to be a substantive epistemological question that is not automatically answered positively by appealing to the conscious character of avowable mental states.

The idea of unarticulated basic self-knowledge may be read instead as a version of the Cartesian thesis of ‘self-intimation’, which says that a subject who is in a mental state is somehow guaranteed to know that she is. I think there are several reasons not to endorse this strong thesis, what with Freudian subconscious mental states and with the discoveries by cognitive scientists of subliminal states guiding our behavior and actions. But even setting these aside, the proponent of self-intimation has to tell us what to make of subjects – such as animals and precognitive children – who *have* mental states, but, due to conceptual limitations, are incapable of passing any judgment on their mental states. It seems highly implausible to maintain that only a creature capable of *making judgments* about its present mental states should be capable of *having* mental states. But it also seems problematic to stipulate that the thesis of self-intimation only applies in the case of creatures capable of self-reflection. Why should having the relevant conceptual wherewithal bring in its train a guarantee that when one is in M one will know it – i.e., have a true and warranted belief that one is in M?

The natural suggestion at this point would be that simply *being* in a conscious mental state and being capable of self-reflection is not sufficient by itself for having self-knowledge; one must also attend to one’s present mental state and actually *exercise* one’s capacity for self-reflection. This seems plausible enough, but if this is what the thesis of self-intimation amounts to, it’s unclear how it can serve to explicate the idea of

unarticulated self-knowledge. For, to require that one actually exercise one's capacity for self-reflection on the occasion of being in M, if one is to have knowledge that one is in M, is to require that, well, one have the occurrent thought that one is in M. It is to require, in other words, that one actually ascribe being in M to oneself in thought. Assuming such an act of self-ascription is not made on the basis of any behavioral evidence, inference, etc., it will readily qualify as an act of avowing by the Neo-Expressivist's lights, provided that it constitutes an act in which the avower gives direct expression to M itself. Self-ascriptions issued upon attending to one's present state of mind are naturally regarded in this way. (Attending need not be construed as an epistemic method of obtaining information or discovery; it can be seen instead as a psychological device for putting oneself in a position to *speak directly from* one's condition. In the special case of avowing in thought, there is, of course, no speaking. Still, there is an issuing of a self-ascription—some mental tokening in which one thinks *that* she herself is in M. Now, on my account, avowals in thought, like avowals in speech, can be seen to enjoy distinctive security: we would not expect any reasons or justification for such avowals, beyond the subject's simply being in the self-ascribed state. And, to take someone to be avowing M in thought is to take her to be expressing (to herself) her M, which means presuming that her avowal is true. Moreover, assuming the avowal is indeed true, we would take the subject who avows to have privileged knowledge that she is in the state she thinks she is in.

What about Rorty's dilemma? If the Neo-Expressivist account of 1st-person authority is right, we may have the materials for recovering the commonsense separation of mind and body *without* resorting to Cartesian Dualism. The states with respect to which we enjoy 1st-person authority, as commonsense would have it, are states that are *expressible* – states that we can *speak from* as well as show through non-verbal expressive behavior. Far from committing us to an ontology of hidden states that are only directly observable by subjects and inferrable by their observers, the commonsense picture is one according to which, as observers, we can see subjects' mental states in their expressive behavior, avowals included. What subjects can do that as their observers we cannot is express the subjects' states. Herein lies their 1st-person privilege.

