

Knowing Selves: Expression, Truth, and Knowledge

Dorit Bar-On and Douglas C. Long, UNC-Chapel Hill

1. Introduction: Avowals and Self-Knowledge

Self-knowledge can be gained in a variety of ways. People often learn things about themselves by observing how they behave in various situations. These situations might be very ordinary ones, or stressful ones, or even those contrived by psychological experimenters. One can also gain self-knowledge by talking with other people and seeing oneself reflected in their comments and attitudes. For example, I may not be aware of how self-centered I am until someone else points it out to me. I then have self-knowledge or self-understanding that I did not possess before. Discussing issues with a trained therapist is supposed to be an especially helpful way to learn about oneself in this way. Much of this knowledge about oneself could be acquired in the same way by others, since they too can observe, or gain access to, what another person says and does.

However, the traditional concept of "privileged access" focuses on the possibility of a very different avenue to self-knowledge, one that is especially secure and uniquely limited to the first-person perspective. Typically this type of self-knowledge is thought to be gained by a kind of "inner perception" called "introspection." I am aware of my thoughts, feelings, and sensations in a direct and unmediated way merely because they occur in me. Others cannot have that sort of access to my mental states. They can, of course, have similar direct access to their own. According to this view each of us can form judgments about our own thoughts and feelings that we articulate either for ourselves, often silently, or for the benefit of others in what we may call "avowals." Such avowals are thought to capture a special form of self-knowledge that has private origins unlike the more public varieties mentioned above.

In our discussion that follows we take issue with the privileged-access model in both its traditional and its more contemporary versions. We present an account of avowals that takes them to be direct expressions of our mental states rather than the products of a special first-person epistemology. This approach calls into question the appropriateness of the phrase "privileged access" insofar as it suggests that the access is epistemic. But it also might be thought to call into question the legitimacy of regarding our avowals as expressing something properly called "self-knowledge." For instance, if avowals are taken to be expressive in the way that nonverbal behavior can be expressive of mental states, one might understandably doubt that they represent suitably justified beliefs. Natural expressions of anger and pain, for example, seem peculiarly noncognitive and unsuited to an account of self-knowledge. We argue, on the contrary, that our particular "Neo Expressive" view can indeed accommodate the idea that avowals represent a special type of self-knowledge.

As we are thinking of them, avowals form a special sub-class of self-ascriptions that can be made in either speech or thought. Roughly speaking, these are present-tense mental self-ascriptions such as: "I am so thirsty," "It feels hot to me in here," "I am scared of that dog," "I am worried about my mom," and "I hope she arrives soon," that are not the products of any inference, and are not made on the basis of any evidence or observation of one's behavior. Such self-ascriptions are often taken to enjoy a special security. Our focus will be on the question whether avowals can be properly said to represent a special kind of knowledge that we have of ourselves, given an expressive account of the source of this security. We will not be assuming that everything we know about ourselves must be explicitly avowed in speech or in thought. Rather, we will be directly concerned with whether what we do, or can, articulate in avowals constitutes a special kind of knowledge.¹

Semantically speaking, avowals appear rather mundane. On their face, they concern contingent matters of fact; that is, they ascribe contingent happenings to particular individuals at particular times. My avowal, “I’m feeling thirsty,” has the same content as any ascription that identifies that same individual, namely me, and ascribes to her the same property at the same time (e.g., “*She* is feeling thirsty” said of me, or “*DB* felt thirsty yesterday” said tomorrow.) You could also use it as a premise in boring logical inferences, such as “I’m also feeling thirsty; so that makes two of us.” This is what we call Semantic Continuity: the claim that avowals are continuous in semantic content and logical structure with ordinary, empirical statements.

Epistemically speaking, however, avowals appear rather extraordinary. They seem epistemically “groundless,” inasmuch as they do not seem to be products of reflection, discovery, examination of evidence, or even ordinary observation or perception. We do not expect an avowal to be backed up by reasons or justification. At the same time, avowals seem highly secure. Unless we suspect insincerity, the fact that someone has made an avowal is as good a basis as any for accepting what the avowal says; and we would normally be hard pressed to challenge or correct an avowal. Avowals thus exhibit notable epistemic asymmetries with a wide variety of ordinary empirical ascriptions. This is the claim we call Epistemic Asymmetry.

An adequate account of the security of avowals should respect both Semantic Continuity and Epistemic Asymmetry. Yet this proves more difficult to do than it may at first seem. The familiar Cartesian view appears to combine Semantic Continuity and Epistemic Asymmetry by invoking a special faculty of introspection that gives each of us immediate, infallible, and therefore logically privileged access to our own present states of mind.² But this view comes at a high price; for it requires us to accept the notion of immaterial particulars that enjoy objective existence, but whose contingent states are infallibly knowable—a very problematic notion at best.

One non-Cartesian way to capture the striking epistemic contrasts between avowals and other ascriptions involves claiming that, surface appearances to the contrary, mental predicates such as “feeling thirsty” or “finding the noise annoying,” have different application conditions in the first- and third-person uses. On an extreme version of this view, which is sometimes attributed to Wittgenstein (erroneously, we believe), avowals are not ascriptions at all. They are simply glorified replacements of natural expressions. Whereas when you say of me, “She has a toothache,” you are making a genuine ascription of a state to a particular individual, when I utter, “I have a toothache,” I do something that is very similar to wincing and holding my cheek. For that reason, my utterance is protected from being false. However, this is only because it cannot be true, either. It is not a truth-assessable ascription in the first place. Consequently, avowals cannot share truth-conditions with other utterances and cannot figure in logical inferences. This means that, if this “Simple Expressive Account” (as we shall call it)³ secures Epistemic Asymmetry, it is only at the cost of flouting Semantic Continuity. Furthermore, it should be noted that the Simple Expressive account would allow us to understand why it is out of place to ask *how* I know, say, that I have a toothache only by denying altogether that my avowal is even a candidate for any kind of genuine self-knowledge.

The dualist and the Simple Expressive views both rely on the following presupposition:

If I can be said to have privileged knowledge that I am in a certain state of mind, then this knowledge must have some *distinct epistemic basis*; there must be a special epistemic method or route that I use to come by this knowledge.

The Cartesian takes it as non-negotiable that we do have such self-knowledge. Impressed by the secure character of this knowledge, he searches for a special, infallible mode of epistemic access that can serve to ground it. The proponent of the Simple Expressive view, on the other hand, seizes on the fact that questions about how I know my own

present mind, and whether I could be wrong, seem out of place, and ends up denying that avowals can be candidates for knowledge any more than natural expressions. As we shall see in Section 2, the “Distinct Epistemic Basis” presupposition is inherited by contemporary successors of the Cartesian and the Simple Expressive accounts. Later we shall argue that this presupposition ought to be rejected.

As a preliminary, we propose to separate two questions, which have not been properly separated in treatments of avowals and self-knowledge:⁴

- (i) What accounts for the *unparalleled security of avowals*? Specifically, why is it that avowals are so rarely questioned or corrected, and are generally so resistant to ordinary epistemic assessments?

And

- (ii) Do avowals serve to articulate *privileged self-knowledge*? If so, what qualifies avowals as articles of knowledge at all, and what is the source of the privileged status of this knowledge?

Question (i) concerns the seemingly unique security of a certain class of performances (in speech or in thought), namely, avowals. It need not be assumed that the only legitimate account of avowals’ security is an epistemic one—i.e., an account that invokes an especially secure way of knowing, or an especially secure epistemic basis on which avowals are made. Indeed, the Simple Expressive account mentioned earlier explains avowals’ security, and thus answers (i) in a way that is clearly non-epistemic in the present sense. By contrast, answering question (ii) directly calls for an epistemological investigation as to what—if anything—qualifies avowals to represent genuine knowledge, and what—if anything—renders this knowledge privileged.⁵

Now, in the case of the Simple Expressive account, it is clear that the non-epistemic, “grammatical” answer it offers to question (i) is coupled with a deflationary answer to question (ii); for the account simply denies that we possess a privileged kind of self-knowledge articulated by avowals. In Section 2 below, we shall consider a more sophisticated grammatical account of avowals that is also deflationary regarding self-

knowledge. In Section 3, we shall sketch our preferred Neo Expressive account of avowals' security, which also provides a non-epistemic answer to question (i). Like the Simple Expressive account, the account we present attributes the special security of avowals to the fact that they do *not* rely on any epistemic access or route. But, unlike the Simple Expressive account, ours does not require deflating self-knowledge. On our view, although the special security of avowals is not due to an especially secure way of knowing, this does not mean we do not possess knowledge—even privileged knowledge—of our present states of mind that can be represented by avowals. If this seems paradoxical, it is due to the unquestioned acceptance of the “Distinct Epistemic Basis” presupposition mentioned earlier: that where there is special knowledge, there must be an especially secure epistemic method of obtaining it, so that in the absence of such a method, there could be no privileged knowledge for avowals to articulate. Once this presupposition is rejected, it becomes possible to combine a non-epistemic account of avowals' security, such as ours, with a non-deflationary view of self-knowledge. In the remaining sections of the paper, we will explain how the proponent of the Neo Expressive account can uphold such a view.

(2) The Special Security of Avowals: Epistemic and Non-Epistemic Views

Let us return for a moment to the Cartesian dualist account and the Simple Expressive account mentioned earlier. We can see the Cartesian account as beginning with question (ii), which is concerned directly with the status of self-knowledge. Having identified a special epistemic route—a privileged access—we each possess to our present states of mind, the Cartesian can go on to represent avowals as articulating self-judgments that are upshots of exercising this special way of knowing, thereby offering an epistemic answer to (i). The Simple Expressive account, by contrast, can be seen as aiming, in the first place, to explain the special security of avowals, as a class of performances. Given the non-epistemic explanation it offers for that security, the Simple

Expressive account seems forced to give a negative answer to question (ii), and a deflationary view of self-knowledge.

Interestingly, both accounts appear to falsify the character of the special security of avowals. On the Cartesian view, avowals enjoy an absolute guarantee of truth, because they are arrived at through the exercise of an infallible form of access. On the Simple Expressive view it also turns out that avowals cannot be false, because, like natural expressions of mental states, they cannot be true, either. Now, we do not believe there can be a logical guarantee of truth for avowals; and we do not believe in their absolute incorrigibility. We think that views that endorse the logical infallibility or incorrigibility of avowals fail to capture the character of avowals, as well as our ordinary treatment of them.⁶ This furnishes us with one good reason (though not the only one) to reject both the Cartesian and the Simple Expressive answers to question (i).

Before turning to the answer we ourselves advocate, we want to examine briefly two other answers found in recent discussions of self-knowledge that can be regarded as successors to the Cartesian and the Simple Expressive views, respectively. The interest these more recent answers have for our purposes lies in the fact that they reject the incorrigibility claim shared by the Cartesian and the Simple Expressive views. First, as a contemporary successor to the Cartesian view, consider the familiar Materialist Introspectionist view. Briefly, according to this view, each of us does enjoy a special access route to our own present states of mind, though not the logically privileged access postulated by the Cartesian. Our special access is causally underwritten, by an in-principle fallible, though highly reliable mechanism, which we human beings possess, as a matter of biological fact. This mechanism is designed to scan, or track our first-order mental states, which, as it happens, are internal states of our bodies (more specifically, our brains). Avowals, in turn, serve to articulate the higher-order beliefs or judgments that are delivered by this tracking mechanism (Armstrong 1968, 94-95).

The Materialist Introspectionist account, like its Cartesian predecessor, portrays avowals as sharing semantic content with third-person ascriptions. It also offers an epistemic answer to the question about avowals' special security (our question (i)). For it explains the special security of avowals as due to the fact that they are reached through an epistemic route that is more secure than the routes available to others.⁷ It thus promises to explain Epistemic Asymmetry without flouting Semantic Continuity. At the same time, the Materialist Introspectionist account avoids the ontological excesses of the Cartesian account, as well as the unrealistic claim that avowals enjoy logical infallibility or incorrigibility.

However, the Materialist Introspectionist account does not capture all aspects of Epistemic Asymmetry. The Materialist Introspectionist offers up the model of self-ascriptions that are upshots of relatively causally direct and reliable channels designed to provide one with information about certain of their own bodily states. This model helps capture the non-evidential character of avowals. But it is ill-suited to capture any principled contrast between avowals and nonmental self-ascriptions that are also non-evidential—e.g., “My legs are crossed” or “I am sitting down,” as made in the normal way. The Materialist Introspectionist must deny that there is a genuine asymmetry between avowals and such nonmental self-ascriptions. (See below, Section 3.) Furthermore, it may be objected that the Materialist Introspectionist fails to capture adequately the *first-person* character of avowals' security. Even if avowals are not guaranteed to be true, it seems crucial to their security that they embody ascriptions of mental states that one makes to oneself. If so, the special security enjoyed by avowals could not transfer to anyone else's pronouncements on my mental states. Yet it is a consequence of the Materialist Introspectionist account of avowals' security that, if it were causally possible for someone else to produce non-evidential ascriptions of present mental states to me (say, through telepathy, or shared brain wiring), their

pronouncements would enjoy as much security as my own avowals. (See Wright 1998, 24). The Materialist Introspectionist account also allows for the possibility of a subject who goes wrong in all of her avowals (say, if her tracking mechanism is faulty). But, even if we allow that a person's avowals can on occasion be false, avowals' security seems in a certain sense "inalienable." As Wright puts it, "[t]here is no such thing as showing oneself chronically unreliable in" one's avowals (1998, 17). Of course, denying these (alleged) asymmetries may well be part of the Introspectionist Materialist view. Our point is just that this involves compromising Epistemic Asymmetry in its full scope.⁸

If one believes that avowals enjoy a distinctive security that cannot be shared by other ascriptions, one has reason to reject non-Cartesian accounts that tie the security of avowals to their contingent epistemic pedigree. The Simple Expressive account mentioned earlier tries to preserve the relatively non-negotiable character of avowals' security by drawing attention to their "grammar" (in Wittgenstein's sense) rather than their epistemology. By drawing a close comparison between avowals and natural expressions, and denying that avowals are reports on present goings-on, the Simple Expressive account is able to portray avowals as protected from both epistemic criticism and correction by others, in a way no other type of ascription would be. But we have seen that the Simple Expressive account compromises Semantic Continuity.⁹

Crispin Wright has offered on Wittgenstein's behalf a view that is designed to avoid this shortcoming of the Simple Expressive account (Wright 1998, secs. X and XI). On this view, which Wright dubs the "Default View," the practice of psychological ascription provides a kind of "bedrock" in accounting for asymmetries between avowals and other ascriptions. The presumed truth of avowals and the fact that they are not subject to ordinary epistemic assessment simply reflect features that are "primitively constitutive" of psychological discourse. "[I]t is simply basic to the competent ascription of the attitudes that, in the absence of good reason to the contrary, one must

accord correctness to what a subject is willing to avow, and limit one's ascriptions to her to those she is willing to avow" (Wright 1998, 41). This default status of avowals is part of what fixes our *concepts* of mental states; it is not a product of non-conceptual facts about subjects' mental life or their epistemic relationships to their mental states. The status is captured by the following constraint: "unless you can show how to make better sense of her by overriding or going beyond it, [a subject's] active self-conception, as manifest in what she is willing to avow, must be deferred to" (Wright 1998, 41).

On the Default View, as on the Simple Expressive view, the special security enjoyed by avowals is a consequence of "grammar," rather than epistemology. But the Default View seems better equipped than the Simple Expressive account to accommodate the semantic continuities between avowals and other ascriptions. For, on the Default View, avowals do involve ascriptions of present mental states to oneself that could equally be made by others. That is, we can take it that my self-ascription, "I am in pain," and your ascription to me, "She is in pain," both ascribe a certain state to the same individual (me). Since they both concern the same mental concept, they are both subject to the "primitive constraint" of "default first-person authority." It is just that the consequences of the constraint will be different for the two ascriptions: the self-ascription, in contrast to ascription by another, will be assigned default correctness. However, the Default View—in contrast to the Materialist Introspectionist view—sharply sets apart mentalistic ascriptions from other ascriptions. Indeed, the view seems to go too far in that direction. For, as Wright presents it, the view builds the (default) correctness of first-person ascriptions into the *truth-conditions* of mentalistic ascriptions "primitively." So the truth of mental ascriptions, unlike the truth of nonmental ascriptions, turns out to be strongly dependent on self-judgments of subjects. Our practices of mentalistic ascriptions are rendered special through being "not accountable to any reality" (Wright 1998, 39).

Even if the Default View were able to adhere to Semantic Continuity, it is not clear how well it captures Epistemic Asymmetry. Note that the default assumption that accords special authority to present tense mental self-ascriptions constrains mentalistic discourse as a whole. The problem is that, by itself, the constraint does not reveal in what ways avowals differ from other mentalistic ascriptions—what makes (some) applications of mentalistic concepts in one’s own case special. A global constraint that applies to *all* mentalistic ascriptions is perhaps apt to explain how mentalistic discourse is different from non-mentalistic discourse. But it is not apt to explain the contrast between first-person and third-person mentalistic ascriptions—a contrast that arises *within* mentalistic discourse itself.¹⁰

Presumably, it is part of the Default View that there is nothing *to* explain concerning the first-person/third-person asymmetries; these are simply constitutive of mentalistic discourse. It is in this sense that the view offers a purely “grammatical,” non-epistemic answer to our question (i) about the special security of avowals. But note that the Default View’s answer to (i), like that of the Simple Expressive account, requires adopting a negative answer to our question (ii) and a thoroughgoing deflationary view of self-knowledge. To accept the Default View is to deny that avowals are apt to represent a special kind of knowledge deserving the epithet “privileged self-knowledge” that can be articulated by avowals. As Wright himself remarks in criticism of the view, adopting the Default View may be nothing more than “merely an unphilosophical turning of the back” on issues surrounding avowals and self-knowledge (1998, 41).¹¹

Our discussion so far suggests that an adequate account of the security of avowals should satisfy the following desiderata:

- (a) The account should explain what renders avowals protected from ordinary epistemic assessments (including requests for reasons, challenges to their truth, simple correction, etc.).

- (b) It should explain why avowals' security is unparalleled: why there are asymmetries in security between avowals and all other empirical ascriptions, including (truth-equivalent) third-person ascriptions and nonmental first-person ascriptions.
- (c) It should explain the non-negotiable character of the security – the fact that it is 'non-transferable' and 'inalienable'.

(Meeting (a)-(c) would amount to explaining Epistemic Asymmetry in its full scope.)

- (d) It should accommodate the continuities in semantic and logical structure between avowals and other ascriptions. In particular, it should present avowals as truth-assessable.

(Meeting (d) would amount to accommodating Semantic Continuity.)

- (e) It should avoid portraying avowals as absolutely infallible or incorrigible.
- (f) It should avoid relying on Cartesian dualist ontology.
- (g) It should allow for the possibility that avowals represent privileged self-knowledge.

(Meeting (g) would afford a non-deflationary answer to our question (ii).)

“Grammatical” accounts of the security of avowals seem potentially better placed than non-Cartesian epistemic accounts to meet desiderata (a) - (c). Non-Cartesian epistemic accounts, on the other hand, seem better placed to meet (d) - (g). None of the accounts canvassed so far meets all the above desiderata. In the remainder of the paper, we sketch an account that, we believe, does so. In the next section, we present our preferred account of avowals' security in a way that brings out its potential for meeting desiderata (a) – (f). Sections 4 and 5 outline some non-deflationary views of self-knowledge that are compatible with our account, thereby showing that it can meet (g).

3. The Special Security of Avowals: A Neo Expressive Account¹²

Consider present-tense proprioceptive reports such as “My legs are crossed,” and self-ascriptions of position (e.g., “I am standing up”) and orientation (e.g., “I am in the

middle of the room”). In the normal case, these nonmental self-ascriptions, much like avowals, are not based on evidence, inference, or ordinary observation. More crucially, as with spontaneous mental self-ascriptions such as “I’m feeling thirsty” or “I think I need to lie down,” these bodily self-ascriptions do not require the self-ascriber to ascribe the relevant property to herself based on recognition of some individual as herself.

Except under unusual circumstances, it would be very odd to raise the question

“*Someone*’s legs are crossed, but is it me?” or “*Someone* is standing up, but is it me?”

According to Sydney Shoemaker and Gareth Evans the reason why this is odd is that in making a self-ascription such as “I am sitting on a chair,” I do not rely on any substantive identification of myself.¹³ I have no reason, grounds, or basis for judging that *someone* is sitting on a chair that are separate from whatever grounds my judgment that I am the one sitting on a chair. When I make such an ascription, I cannot be in error as to the identity of the individual to whom I ascribe a particular property, though I may well be in error regarding *what* I ascribe to myself. Such bodily self-ascriptions are, in the terminology of Shoemaker and Evans, *immune to error through misidentification* (IETM, for short).¹⁴

We should note, first, that unlike certain Wittgensteinian proposals, Shoemaker and Evans’ diagnosis does not involve denying that the pronoun “I” can genuinely refer to a particular individual.¹⁵ Instead, it denies that reference always requires an act of recognition, or “thick” identification (as we might put it) on the part of the user of the referring expression.¹⁶ My ascription “I am sitting on a chair” is semantically about a particular individual, namely myself, even though, under normal circumstances, I do not do anything to ascertain that I am the “right” subject to whom to ascribe the property of sitting on a chair. Secondly, on the proposed analysis, ascriptions are IETM only insofar as they are arrived at in a certain way. The very same ascription can be IETM or not, depending on whether it relies on an identification of the ascription’s subject. If I say,

“My legs are crossed,” upon looking in the mirror, my ascription will be vulnerable to errors of misidentification, since it relies on an identification judgment that may be false (say, if the person in the mirror is someone else). Finally, note that the security afforded by ascriptions that are IETM is negative. What renders a self-ascription IETM is not that the self-ascriber utilizes an especially secure way of *identifying* herself, but rather that it does *not* rely on any particular epistemic way of identifying the subject of the ascription.

Avowals represent one paradigm of ascriptions that are IETM, but their security goes beyond this immunity.¹⁷ To see this, consider again reports such as “I am raising my arm” or “I am spinning around the table,” which are IETM (when made in the normal way). Such first-person reports are rendered on a different, and causally more direct, basis than are their third-person analogues. Yet they are completely and straightforwardly open to brute error and subject to correction. My proprioceptive or kinesthetic abilities may become impaired, so that, looking at me, you could simply see that I am not raising my arm, or that I am standing still, my own pronouncements to the contrary notwithstanding. And your verdict will typically carry more weight than mine, even though it is rendered on the basis of sense perception, which seems causally less direct than proprioception or our kinesthetic sense. By contrast, avowals are not straightforwardly open to correction simply on the strength of an observer’s contrary judgment. And it is not easy to imagine anything coming to play a role analogous to that played by your direct observation to the effect that, e.g., I’m not raising my arm. Although avowals are not guaranteed to be true and are not absolutely incorrigible, they do not seem open to brute epistemic error, and are much more resistant to doubt, epistemic criticism, and correction than nonmental ascriptions that are IETM.¹⁸

Notice that, in the normal case, as I say or think, “I am feeling thirsty,” it would be as out of place to ask whether it is thirst that I am feeling as it would be to question whether it is I who is feeling the thirst. As I simply avow being thirsty, as opposed to

conjecturing about my own state of mind on some basis, the character of the state I avow is no more in question for me than the identity of the person avowing. The same applies for the content part of avowals of propositional attitudes, such as “I am wondering *whether it’s time to leave.*” The point here is not to deny that one may be unclear about one’s present state of mind. (Notice that such unclarity may itself be revealed in an avowal, viz., “I am feeling something, but I am not sure what it is.”) The point is rather that, from the perspective of the avowing subject, the ascriptive part of the avowal is no more open to question than is the subject part. In view of this, we may propose that:

When avowing, a self-ascriber is immune not only to errors of misidentification—i.e., errors about who is the subject of the ascription; she is also immune to errors of misascription—i.e., errors concerning what is being ascribed.

The suggestion is that this additional *ascriptive immunity* can serve to mark an important contrast between avowals and nonmental “I”-ascriptions that are IETM. Understanding the source of ascriptive immunity can illuminate the distinctive security of avowals.

Immunity to error, as we are thinking of it, is essentially a negative form of security. It comes from not making a certain kind of judgment, as opposed to being especially good at making such judgments. This is one grain of truth in the Simple Expressive account mentioned earlier, which rightly connects the security of avowals to the absence of “epistemic targeting” on the part of the avowing subject. However, we also pointed out that the Simple Expressive account excludes avowals from the category of true or false ascriptions, thereby violating Semantic Continuity. We now want to explain how the account we favor is not subject to the same complaint.

The key point to appreciate is that immunity to error as understood here pertains to the epistemology of avowals, not to their semantics. To say that avowals are immune to errors of misascription is to point out that avowing, as such, does not involve “trying to get it right” regarding one’s present mental state or making an epistemic effort, in the sense of putting to use some epistemic method, or mechanism, or having an epistemic

basis, etc. As with immunity to error through misidentification, to say that avowals enjoy ascriptive immunity is only to deny that they are subject to certain kinds of *epistemic errors*, by which we mean mistakes of confusing one state (or content) with another, and other failures involved in ascertaining the presence or character of one's present mental states. The distinctive security of avowals is thus portrayed not as due to a guarantee of epistemic success, but rather as due to the fact that there is no room for epistemic failure. This is because to avow, "I've got a splitting headache" or "I'm thinking it's time to start," is *not* to make a report about one's current mental state on this or that epistemic basis. But from that it does not follow that avowing cannot involve making a genuine mentalistic ascription to oneself. The entailment would go through only if saying or thinking something with a certain semantic content—in particular, issuing a genuine ascription of a state to an individual—always required deploying some robust epistemic means for determining that one is in the relevant state, as well as for picking oneself out. But this is not so.

The idea here can be summarized by the following slogan: *semantic achievement need not be epistemically underwritten*. Just as one can pick oneself out as the subject of ascription without using special epistemic means of recognizing oneself, so one can make a genuine ascription of a mental state to oneself without employing a special epistemic route to that state. To see how this can work, reflect for a moment about the following special case of avowing: "I am thinking that it's time to move on." Such an avowal seems maximally secure. In making it, I cannot misidentify the subject of my ascription; I cannot fail to ascribe to myself some thought or other that is crossing my mind; and I cannot fail to ascribe to myself the thought *that it's time to move on*. Indeed, such an avowal is self-verifying – it makes itself true.¹⁹ Contrast this avowal with the self-ascription: "I am thinking something disturbing," which is clearly not self-verifying. What makes the avowal "I am thinking that it's time to move on" self-verifying is the fact

that the very act of avowing the thought inevitably involves the thought “passing through” my mind, thereby rendering my ascription true.

The peculiar security of avowals of entertained thoughts seems connected with the fact that the thought’s content is directly spelled out in the very act of avowing. This is a feature that is shared by *all* intentional avowals. A typical avowal of a hope will involve going through the relevant state’s propositional content: “I hope *she gets here on time*.” A typical avowal of a desire will involve mention of its intentional object: “I want *a cup of tea*.” Intentional avowals thus contrast with oblique self-ascriptions such as, “I am hoping for the same thing I did yesterday” or “I want the thing that you want.” In intentional avowals, the assignment of intentional content is not achieved by ascertaining in some indirect way that the relevant content is the right content to assign to one’s state. This is why there is no room for mistaking what content to assign to the state ascribed by an avowal. In avowing, content is assigned through an explicit articulation of it in the very act of avowing. Explicit articulation obviates the need for epistemic mediation in the assignment of intentional content.

Self-verifying avowals such as, “I am thinking that it’s time to move on” present a special limit case. When the state ascribed is simply one of entertaining a certain thought, all I need to do in order to succeed in *truly* ascribing to myself the entertaining of the thought *that p* is to entertain the thought *that p*, which I am bound to do if I ascribe the thought explicitly. When I avow a presently entertained thought, as opposed to ascribing it to someone else, explicit articulation of the thought does more than merely display the content I am prepared to assign to the ascribed state. Rather, it serves directly to express—in the sense of giving voice to—the very condition of entertaining the thought. This is an idea we can seize on: the idea of *ascription through expression*. For, once suitably generalized, it may explain how avowals involve genuine self-ascriptions of contingent present states, even though they do not involve the use of epistemic methods

of identification and recognition. Furthermore, it can help us to see how avowals, though not logically infallible or incorrigible, are nevertheless especially apt to be true.

In his “Language as Thought and Communication,” Sellars (1974) distinguishes three senses of expressing. In the *action* sense, a person intentionally expresses a condition of hers by “venting” it or “giving it voice” (in our terminology). In the *causal* sense, an utterance or piece of behavior expresses an underlying condition by being the culmination of a causal process beginning with that condition. And in the *semantic* sense, a sentence, for example, expresses an abstract thought or judgment by being a (conventional) representation of it. Expressing a hope, say, in the first two senses requires expressing *one’s* hope, whereas expressing a hope in the semantic sense is simply expressing hope, without its necessarily being one’s hope. In this same sense, the linguistic locution “Regrettably *p*” expresses regret, although on a particular occasion, it may not be used for—or succeed in—expressing the utterer’s regret.

The case of self-verifying avowals illustrates the following idea. Spelling out the content of a self-ascribed state, which involves making an ascription that expresses the content in the semantic sense, can amount to *my* expressing the state in the action sense. I give voice to my state of entertaining the thought through my use of an ascription that gives semantic expression to the state’s content. This idea may be generalized beyond the special case of self-verifying avowals, to all intentional avowals. Consider first avowals of beliefs. Saying or thinking, “I believe John is angry with me” is not self-verifying. Still, if this self-ascription is an avowal, the point of making it is to “vent” my (first-order) belief that John is angry with me. (This, by the way, would help illuminate the anomaly of so-called “Moore-propositions.” There is no overt or formal contradiction, for example, in “I believe that John is angry with me, but John is not angry with me.” But seeing the first conjunct as a belief avowal along the lines proposed here allows us to identify an “expressive conflict.” In the first conjunct, I express my belief

that John is angry with me through avowing it, and in the second conjunct I express the contrary belief by sincerely asserting that John is *not* angry with me.²⁰)

More generally still, we should recognize that all intentional avowals, whether of beliefs, or of hopes, desires, angers, fears, and so on, play an expressive role as described above. A typical avowal such as, “I hope John will come,” or “I’d really like some water!” or “I’m mad at you, Mom,” will play a role similar to that played by more direct expressions of one’s own intentional states, such as saying or thinking, “John will come, won’t he?” (hopefully) or “Water, please!” (eagerly) or “Mom!!” (angrily), and so on. In both types of cases, the articulation of the state’s intentional content is in the service of directly expressing one’s intentional state. If so, then we can see why intentional avowals are more secure than intentional ascriptions to others as well as being more secure than “alienated” self-ascriptions of intentional states (e.g., ascribing to myself hatred of my brother on the basis of therapy). On the present proposal, the point of avowing an intentional state is not so much to provide a descriptive report of it as to “vent” it. What is distinctive of acts of avowing an intentional state, as opposed to reporting its presence in oneself or in others, is not that the self-ascription is arrived at on a distinctively secure epistemic basis. Quite the opposite. When avowing an intentional state, a subject expresses her intentional state in the action sense. Such expressive acts do not involve traversing any “epistemic distance,” so to speak. It is for this reason that no space opens up for epistemic errors of misascription.²¹

Intentional avowals in general are not self-verifying. In the special case of an avowal of an entertained thought, the act of articulating the thought *that p*, whether to others or to oneself, whether in language or in some other medium or representation, requires actually being in the state of entertaining that thought. So in this case, expressing *p* in Sellars’ semantic sense suffices for success in expressing *one’s* thought that *p*, thereby issuing a true self-ascription. But there is more to being in a state of

hoping *that p*, wishing *for x*, being afraid *of y*, etc. than the spelling out of the relevant intentional content. So, while the point of articulating the content may still be to express one's first-order intentional state, the articulation does not guarantee that one will succeed in expressing one's hope or wish that *p*, etc. The truth of the avowal is not automatically secured through the act of avowing; it is not self-verifying.

Not all avowals involve an explicit articulation of intentional content. Consider, for example, "I'm so uncomfortable!" or "I'm really thirsty" or "I feel very achy." We think, however, that the expressive idea can be generalized beyond intentional avowals to cover also so-called phenomenal avowals. The proposal is this:

Avowals in general are expressive acts in which the subject "gives voice" to—or speaks directly from—a first-order mental condition.

It is for this reason that avowals enjoy more security than both ordinary empirical reports and all bodily self-ascriptions.

The present proposal bears obvious affinity to the Simple Expressive proposal, according to which avowals, like natural expressions, are expressive of subjects' mental conditions. Notice, however, that in motivating this idea we have followed a very different path from that taken by the more traditional expressive approach. For we have here led up to the relevant similarity by considering first a case as far from moans and groans as seems possible—the case of a fully articulate self-ascription of an entertained thought. The point of doing so was to show how the idea of avowals expressing the states they ascribe can get purchase without an objectionable assimilation of avowals to natural expressions. For, while there are some similarities between avowals and natural expressions, there are also important dissimilarities.

Consider: I see a mean-looking Rottweiler and I am overcome by fear. I may (a) let out a gasp, or (b) exclaim: "That dog is scary!" or (c) say: "I am so scared of that dog!" Semantically speaking, these are three very different performances. The first

involves a natural, non-linguistic expression of my fear. It is linguistically inarticulate; it makes no reference to dogs, or being scared, or anything. The other two involve linguistic utterances that say quite different things. (b) attributes to the dog the quality of being scary, whereas (c) says that someone is in the condition of being scared of the dog. Still, in a given context, the three performances can be on a par, epistemically speaking. In particular, though the self-ascriptive utterance ascribes to me a certain state, it need not involve my epistemic determination that I am in that state. It can simply be my way of giving direct expression to the state. This makes for an epistemic similarity between avowals and natural expressions, and for epistemic asymmetries between avowals and other ascriptions, including mental self-ascriptions that are made on the basis of evidence or inference.

Avowals are similar to natural expressions in terms of the *act* performed in issuing them. However, they use quite dissimilar *expressive vehicles*. Unlike acts of natural expression, avowals as such do not use expressive means that display or exhibit the mental states. Avowals are *self-ascriptive expressions* of one's present first-order mental states that give the states articulate voice. An avowal of joy does not *show* the joy as might a cheerful voice or hug. Rather, it *tells* us something. For all that, sometimes a telling can replace, or play the role of, a showing. To use previous terminology, in both cases the expressive vehicle—the articulate utterance in the one case, the hug in the other—serves to express the subject's first-order condition in the causal sense, and the subject can be said to express that condition in the action sense. But only in the case of avowal does the vehicle also give semantic expression to the claim that the subject is in such and such condition.

To summarize, on the Neo Expressive account, an avowing subject speaks from a first-order condition. She is immune to epistemic errors of misascription, as well misidentification, because there is no epistemic distance between her and the condition

she speaks from. In this, avowals resemble natural expressions and contrast with all other empirical ascriptions. So Epistemic Asymmetry is fully respected. However, the ascriptive immunity of avowals is not the immunity of grunts and winces, which prevents them from being true *or* false. Avowals issue in semantically articulate, true (or false) mental self-ascriptions, which can figure in inferences, and can be interchangeable with other ascriptions that have the same truth-conditions. So Semantic Continuity is preserved. The combination is achieved by recognizing that, in avowing, I express my mental condition not by showing it, but rather by calling its name, as it were, and by spelling out its intentional content when it has one. I express my state of mind by *speaking* my mind (which, N.B., is something I can do out loud, in public, as well as in *sotto voce*, with no audience in mind.)

It is often objected that, if avowals are not seen as assertoric reports, then avowals *cannot* be regarded as true or false ascriptions. *Ipsa facto* they cannot be things subjects can be said to *know*, let alone know with any special authority. We hope it is by now clear how an expressive view can accommodate the truth assessability of avowals. It is helpful to distinguish in this context the *products* of avowals—namely, articulate self-ascriptions—from the *process* involved in producing them. The fact that avowals serve to express a subject’s first-order mental state, and not her higher-order judgment about its presence and character, is not a fact about the grammar or semantic content of avowals’ products; rather it has to do with the process involved in producing them. An utterance or thought can have the content of a self-ascription in that it can be true if and only if a particular subject (the avower) is in the state referred to by the relevant mental term. Yet it may not be produced by a process that leads to the subject’s *judging that* she is in that state.

An analogy may help to make the point. Suppose I utter: “I promise to take you to dinner tonight.” I am using a sentence that says that D promises to take [so-and-so] to

dinner [on a certain evening]. But, in the typical case, I will not be reporting a promise; I will be making it. Similarly, if I utter: “I am really thirsty,” as I desperately reach for water, I will be using a sentence that is true under the same conditions as your observation: “You are really thirsty.” Still, in uttering the sentence I may not be reporting my thirst but simply expressing it.²²

It should also be clear why the self-ascriptions that are the products of avowals are not merely true or false, but are especially apt to be true. For, insofar as the products of avowals are seen as “coming from” subjects’ conditions as do natural expressions, they will share in the immediacy and reliability of natural expressions. If an avowal of a belief that p , or a fear of x , or a pain, is strongly presumed to be true, this is because it can be presumed to take us to the subject’s relevant first-order condition as directly and as immediately as would the subject’s sincere statement that p , his pointing at x while trembling, or his wincing, respectively.²³ No other type of performance exhibits this combination of expressive immediacy and truth assessability. Ordinary empirical reports, whether first- or third-person, though truth-assessable, are epistemically mediated. They rest upon some epistemic basis, and to that extent, they do not enjoy ascriptive immunity to error, even when highly reliable. Acts of natural expression, though epistemically direct, do not have truth-assessable products. Such acts are immune to epistemic errors by simply being excluded from the category of truth—and thus knowledge.²⁴

4. Avowals’ Security and Privileged Self-Knowledge

Let us go back to the two questions we separated in Section 1. We now have an answer to the first question:

- (i) What accounts for the *unparalleled security of avowals*? Specifically, why is it that avowals are so rarely questioned or corrected, and are generally so resistant to ordinary epistemic assessments?

Avowals' unparalleled security, we have suggested, can be plausibly and realistically explained through a unique combination of immunity to epistemic error and special "truth-aptitude" that are a consequence of their expressive character. However, inasmuch as this explanation dissociates the special security of avowals from the employment of a special way of coming to know about one's mental states, it clearly falls short of assuring us that avowals represent *any*—let alone privileged—knowledge. In other words, the non-epistemic answer we have given to (i) does not by itself provide an answer to the second question:

(ii) Do avowals serve to articulate *privileged self-knowledge*? If so, what qualifies avowals as articles of knowledge at all, and what is the source of the privileged status of this knowledge?

Worse still, it may seem that the Neo Expressive account we have offered in answer to question (i) actually precludes giving a positive answer to question (ii) and commits us to a deflationary view of self-knowledge. We will argue that this is not so.

We do not deny that the Neo Expressive account of avowals' security we support can be coupled with a deflationary view of self-knowledge. The Neo Expressive account explains the special security of avowals by portraying them as self-ascriptive expressions that are immune to epistemic errors. Given that they are also especially apt to be true, we can see how others can use avowals as a source of articulate, reliable, and correct information about the self-ascriber's present mental condition. But question (ii) is about *self-knowledge*. Could avowing subjects themselves be credited with privileged knowledge about their present states of mind? One can acknowledge that we have a distinctive ability to produce secure mental self-ascriptions, without thereby accepting that the exercise of this ability is apt to yield any kind of knowledge. For, one can

maintain that possessing genuine knowledge requires more than the production of true self-ascriptions that enjoy expressive security.

Thus, consider one standard and familiar view—the so-called JTB model—according to which knowing *that p* requires having a *justified and true belief that p*. On this view, having self-knowledge in the relevant sense would presumably require having true beliefs about one's present mental states for which one possesses a distinctive kind of epistemic justification. The proponent of the Neo Expressive view faces two challenges. First, to explain whether and how, on that account, avowals are apt to represent *beliefs* that avowing subjects have about themselves, and, second, to explain in what one's (*special*) *epistemic justification* for those beliefs could consist. (So the second task is itself twofold: to explain not only how one could have *any* epistemic justification for the beliefs represented by avowals, but also what makes this justification special.) If the proponent of the Neo Expressive account cannot meet these challenges, then, *assuming the JTB model is accepted*, he would be forced into endorsing a deflationary view of self-knowledge. He would have to acknowledge that, although avowals may be especially secure pronouncements that, unlike natural expressions, are apt to be *true*, they cannot amount to any kind of knowledge, just like natural expressions.

We share misgivings that have been expressed in the literature concerning the correctness of a JTB analysis of knowledge, about both the sufficiency and the necessity of its “belief” and the “justification” conditions.²⁵ Nonetheless we find it useful to refer to that analysis as a benchmark in explaining the relation between the Neo Expressive view of avowals and self-knowledge. For the analysis attempts to capture two requirements on knowledge that seem intuitively correct: first, that the subject who

knows something must somehow be cognitively related to that which she is said to know, and second, that such a subject must not merely happen to believe what is in fact true. In what follows we will examine the extent to which the self-knowledge representable by avowals accords with the traditional JTB model, and consider ways in which avowing subjects can be said to satisfy the intuitive requirements the model attempts to capture.

Let us consider the belief condition first. The Neo Expressive account maintains that, if avowals are distinctively secure, this is because they give voice to subjects' first-order conditions *rather than* to higher-order judgments that they have formed about those conditions. This seems at first glance to tell against the idea that avowing involves a subject's believing that she is in a particular mental state. Since the JTB model requires one to have a belief in order to be said to know, then avowals seem to fail the JTB test for knowledge. Often we come to believe explicitly that *p* by going through an epistemic process of investigation or evidence gathering and inference or observation. In such cases it is appropriate to ask the believer for some sort of justification or reasons supporting the belief. According to the JTB analysis of knowledge, when the justification is sufficient, such belief, if true, counts as knowledge.

Some philosophers have resisted the JTB analysis on the grounds that belief is not a core element of knowledge. Zeno Vendler, for instance, has argued that having knowledge actually contrasts with believing. This is because to know that *p* is to be in robust "epistemic contact" with the *fact* that *p*, or with the state of affairs *p* represents; whereas to believe that *p* is to be in a certain relation, perhaps psychological or cognitive, to the *proposition* that *p* (Vendler 1972, 138-49). Timothy Williamson, while not endorsing Vendler's account of the contrast, argues that knowing (like seeing and

remembering) is, whereas believing is not, a factive verb designating a relation that one can only have to truths (Williamson 1995, 551). Accordingly, we can distinguish two different notions of “belief.” When someone considers evidence and comes to have a belief or forms a judgment that something is the case, she has belief in the “opining” sense. When subjects “opine” that something is the case, it is appropriate to expect that they be able to justify their opinions by offering relevant evidence or reasons. Clearly, this is not, according to our view, a sense of belief that is represented by avowals. It does not seem appropriate to suppose that an avowing subject is of the opinion, e.g., that he is in pain, or that he would like some tea, let alone that his avowal is in need of justification or reasons. However, there is a more minimal notion of belief, according to which a subject believes that p , provided only that he accepts p as true. In this “holding-true” sense as we shall refer to it, if I know that the vase I just dropped is broken, I believe that it is broken, even if it seems inappropriate to suggest that I have come to the opinion that it is broken.²⁶

The Neo Expressive account does not require denying that avowals represent beliefs that subjects have about themselves in the holding-true sense. Recall that, on the Neo Expressive view, avowing involves using an expressive vehicle that has the semantic content of a self-ascription. The product of an avowal says something about the self-ascriber. Suppose I now avow, “I am really thirsty.” The process of issuing the avowal may not involve my forming the belief or judgment that I am thirsty on this or that basis. The self-ascription I produce thus does not express that judgment in the causal sense of “express,” and in producing it I do not express my own judgment to that effect in the action sense of “express.” Still, the avowal’s product is something that semantically

expresses that self-judgment. Even if avowing is not an act of reaching or arriving at a self-judgment, it makes available to me, as well as to others, a claim concerning a present state of mine—a content which, we may assume, I understand perfectly well as I produce the self-ascription. The content avowed can be reasonably regarded as something that I hold to be true. So there is a sense in which I can be said to believe that I am thirsty when I avow, “I am thirsty,” even if the self-ascription I issue is not the product of my having ascertained in some way that I am thirsty. Although in avowing, “I am thirsty,” I may not be opining that I am thirsty, I may still be said to hold it true that I am thirsty.

We submit, then, that there is no reason for the Neo Expressive account to deny that avowals represent beliefs that subjects have about themselves. Subjects can be credited with the relevant beliefs inasmuch as they can be seen as holding true what is semantically expressed by the self-ascriptions they issue when avowing. Thus, maintaining as we do that avowing involves a subject’s expressing her first-order mental condition, rather than her own judgments about those conditions, does not commit us to denying that avowals can represent beliefs in the sense required for knowledge.

However, it may be asked whether the sort of beliefs represented by avowals can be justified in a way that satisfies the JTB account of knowledge. As noted earlier, the JTB model’s justification condition aims to capture a contrast between knowing and merely having a true belief. And so even if we understand belief in the holding-true sense, it will be insisted that knowledge requires more than holding true what is in fact true. Views that take avowals to express beliefs that we form on a secure epistemic basis can easily accommodate the idea that we possess a straightforward epistemic justification for our avowals. By contrast, it may seem that, by the JTB standards, the Neo Expressive

account could at most credit us with beliefs with which we simply find ourselves, or with judgments we are simply unable to deny upon consideration.

Is there then an appropriate notion of justification that is relevant to the case of avowals? Suppose it is held that a knowledge claim can be justified only if the person making the claim has reached a judgment through some epistemic effort or act, such as drawing an inference, judging on some basis, ascertaining in some way, simply observing, or even just attending to the relevant phenomenon. If a person reaches a judgment on such an epistemic basis, then he is justified, because he has reasons for his judgment. The Neo Expressive view excludes such justifications insofar as it does not countenance the epistemic formation of judgments or beliefs. But it seems unreasonable to expect that every knowledge claim should be backed up by justification in this sense. There are other cases of knowledgeable-though-not-explicitly-justified beliefs, such as ordinary memory claims, inference rules used in logical reasoning, and other *apriori* judgments. In addition, it is plausible that we gain knowledge via perception by ascertaining or discovering what is the case and that this process is not properly described as arriving at beliefs (opinions) that require justification. In seeing that the vase is broken, I know that it is. The focus is on my grasp of that fact rather than on reasons for my believing it to be broken. Although in the case of avowing pain, for example, there is no need for me to ascertain or discover that I am in pain, here too we might deny that my knowing the relevant fact must consist in my forming a corresponding belief that requires justification by reasons.²⁷

As pointed out earlier, unless we suspect insincerity, it does not seem appropriate to ask a subject to justify a particular avowal. However, this does not by itself explain

why we should be able to count avowals as representing self-knowledge. It seems incumbent upon us to provide some philosophical account of what renders them a species of self-knowledge, as opposed to mere outbursts that are informative to others.

In Section 1, we called attention to the Distinct Epistemic Basis presupposition that is shared by both deflationary and epistemic-basis accounts of self-knowledge. In its general form, the presupposition states that possessing any distinct kind of knowledge requires making judgments or forming beliefs on some distinct epistemic basis. We can now see this presupposition as simply expressing commitment to a combination of the belief and the justification conditions laid down by the JTB model. We have proposed that there is a sense of “belief”—namely, the holding-true sense—that would allow us to accept the belief condition. We now want to propose, in concession to the justification condition, that knowledge properly so-called may require, in addition to truth, some kind of *epistemic warrant*. However, justification in the narrow sense of relying on an epistemic basis or having reasons is only one kind of epistemic warrant. There may be others. The slack between knowing and holding true what is in fact true may be picked up by something other than justification in the narrow sense. If so, we can reject the Distinct Epistemic Basis presupposition without doing injustice to the spirit of the JTB model.

5. Three Roads to Self-Knowledge

In this section, we sketch three alternative views that would allow us to regard avowals as articulating self-knowledge, even though they do not represent justified beliefs in that narrow sense. We will not aim to develop these views in any detail, or to adjudicate conclusively among them. Our aim will be the modest one of suggesting

avenues open to the Neo Expressive proponent who maintains that avowals represent a kind of knowledge.

5.1 The “Low Road” to Self-Knowledge

According to the epistemological view known as “reliabilism,” beliefs a subject has can constitute knowledge even when they are not justified by reasons the subject has for thinking the beliefs to be true. Rather, they can amount to knowledge “due to their being reliably produced by truth-conducive mechanism or processes, such as memory and perception” where “[r]oughly speaking, a belief-producing mechanism is reliable, or truth-conducive, only if it tends to produce more true beliefs than false beliefs in the relevant situations” (Moser 1986, 5). The reliabilist view does not deny that a subject’s true beliefs must meet a further condition to count as knowledge, but it does reject the “internalist” view that the knowing subject must himself know or be aware of the conditions that allow his beliefs to constitute knowledge. The reliabilist model has seemed especially suited to accommodate the justificatory pedigree of perceptual beliefs in those cases where the subject makes no inference or uses no evidence in reaching the belief. The reliabilist insists that we can possess genuine knowledge of such propositions as “There’s a moose in front of me” or “This dress is red,” even if we do not arrive at such beliefs by inference from more basic justified beliefs, provided we acquire our beliefs through the operation of a distinctively reliable mechanism. The reliability of the mechanism provides the epistemic warrant subjects are said to have for the beliefs that are produced by the mechanism.

This feature of the reliabilist view suggests a possible marriage between the Neo Expressive explanation of the special security of avowals and a reliabilist view of what renders them self-knowledge. On this combination, avowals do not represent beliefs that are arrived at by a subject on the basis of inference or any other form of justification that is available to the subject from an “internal” point of view. Nor is the special security of

avowals to be understood as a consequence of their expressing a subject's own reliably formed beliefs (judgments) about her present states. To explain the security of avowals in that way would be to offer an *epistemic* account, which we have rejected. Rather, the reliabilist position we are now imagining endorses the Neo Expressive idea that avowals express first-order mental conditions and are unmediated by higher-order beliefs. If avowals represent beliefs that subjects have, it is only in the more modest, holding-true sense. Still, avowing issues in truth-assessable self-ascriptions, which do not just happen to correlate well with the mental states to which they refer. They are apt to be true, because the mechanisms underlying expressive behavior are truth-conducive. (The reliabilist might appeal to the reliability of natural expressions as indicators of subjects' mental conditions, and tell a story about how articulate self-ascriptions could inherit that reliability.) Courtesy of the reliable mechanism involved in producing them, avowals can represent self-knowledge.

To clarify: What we are envisaging here is a reliabilist answer to the question "What gives avowals the status of knowledge?" which is part of question (ii). We have rejected the reliabilist answer to question (i), "What makes avowals especially secure?" along with other epistemic answers. Adopting the Neo Expressive account of avowals' security means denying that avowals are mediated by higher-order beliefs that subjects have formed about their first-order mental conditions. *Ipsa facto*, it means that one cannot invoke the reliability of the mechanism by which subjects' higher-order beliefs were arrived at as grounding avowable self-knowledge. Rather, the proposal we are exploring now is that, if avowals are seen as representing beliefs in the relatively thin, holding-true sense, one could regard the reliability of the mechanism(s) underlying self-ascriptive expressive behavior as the source of warrant for what the avowing subject holds true. In holding true what an avowal says, a subject does not merely hold true what

is in fact true. Avowals do not merely happen to be true; they reliably correlate with the mental conditions they ascribe.²⁸

We may not find this stark reliabilist picture very appealing as a basis for an account of self-knowledge. It seems to portray the avowing subject as a kind of input-output device, like a computer that flashes appropriate self-ascriptive messages about its internal processes on a display screen.²⁹ No matter how reliable the subject's self-ascriptive "output," it is not at all clear why we should call it knowledge.³⁰ Avowals might provide the rest of us with information about how the subject feels, perhaps in much the way moans and giggles do. But there does not seem to be enough in the reliabilist answer to question (ii) to reassure us that avowals represent genuine knowledge, properly continuous with other kinds of knowledge. To be sure, an adequate account of self-knowledge that is compatible with the Neo Expressive account should avoid explaining the reliability of avowals by the fact that the subject ascertains what she feels before announcing it. Such an account introduces excess epistemological machinery that puts the avowing subject at one remove from her present mental life, as though she were witnessing her mental states and reporting on them, rather than undergoing them and giving them direct voice. On the other hand, the foregoing expressive-reliabilist account seems to stray too far in the other direction. For it does not preserve the idea that, if avowals represent self-knowledge, it is only because avowing subjects do more than just "erupt" appropriately.

It may be possible for the expressive-reliabilist to go beyond the idea that having avowable self-knowledge is simply a matter of producing reliably correct, semantically articulate signs of one's present states. It is doubtful, however, that the expressive-reliabilist view could ground a distinctive form of knowledge deserving of the epithet "*privileged* self-knowledge." On the reliabilist picture, self-knowledge is entirely optional. Whether, and to what extent, the avowals of any given subject represent self-

knowledge is completely contingent upon whether and to what extent the mechanism producing them is reliable or truth-conducive. Not only is it a biological accident that, in addition to having mental states, we have avowable knowledge of those states, but it is straightforwardly possible for an individual subject to lack such knowledge entirely. Yet our ordinary ascriptions to individuals of avowable self-knowledge do not seem to ride on ascertaining their individual reliability in producing true mental self-ascriptions. Unless linguistic competence or sincerity is in question, we are prepared to credit individuals with avowable knowledge of their basic beliefs, attitudes, feelings and sensations. If the reliability of subjects' self-ascriptions is relevant to avowable self-knowledge, it does not seem to be what makes or breaks it.

Even though one could adopt a reliabilist epistemology of self-knowledge while accepting our (non-reliabilist) Neo Expressive account of avowals' special security, doing so risks distorting the true picture of self-knowledge. We suspect that a reliabilist answer to question (ii), regarding the nature of self-knowledge, would compromise *privileged* self-knowledge, for much the same reason that the reliabilist answer to question (i) would end up compromising Epistemic Asymmetry. (See above, section 2). Just as the reliabilist answer to question (i) can accommodate some of Epistemic Asymmetry, but cannot do justice to its full scope, so the reliabilist answer to question (ii) can perhaps accommodate the existence of avowable self-knowledge, but does not do justice to its privileged status. Still, it is important to recognize that the expressive-reliabilist view represents one avenue potentially open to a proponent of the Neo Expressive account who does not wish to deny that avowals can represent self-knowledge.

5.2 The High Road” to Self-Knowledge

The expressive-reliabilist account of how secure avowals might constitute a form of knowledge builds up self-knowledge from below, adhering to the causal facts about

avowals—specifically, the fact that their production is sustained by reliable mechanisms. It relies on a minimalist analysis of self-knowledge that is designed to be equally applicable to other kinds of knowledge. It thus assures us that there is nothing out of the ordinary about avowable self-knowledge. In this sense, it paves a “low road” to self-knowledge. At the other end of the spectrum of non-introspectionist views of self-knowledge lies a “high road” approach, according to which avowals represent genuine knowledge that, additionally, enjoys a unique status. Whereas on the low-road approach, possessing avowable self-knowledge is entirely optional, on the high-road approach it can be argued on apriori grounds that avowable self-knowledge is obligatory.³¹

As suggested above, one can be epistemically warranted in one’s claim or belief even if one is not justified in the narrow sense—even if one has not engaged in any epistemic effort to form an opinion about the relevant matter, and even if one cannot provide any reasons for what she says or thinks. In “Our Entitlement to Self-Knowledge,” Tyler Burge says, “I take the notion of epistemic warrant to be broader than the ordinary notion of justification. An individual’s epistemic warrant may ... also be an entitlement that consists in a status of operating in an appropriate way in accord with norms of reason, even when these norms cannot be articulated by the individual who has this status. We have an entitlement to certain beliefs or to certain logical inferences even though we lack reasons or justifications for them.” (1996, 3)³² Burge’s point is illuminated by a familiar analogy from the legal domain. One can be legally entitled to something without engaging in any specific action or effort designed to secure her legal hold. One can be legally entitled to do something, for example, provided only that in doing it she acts in accordance with applicable legal norms, regardless of whether she is aware of those norms. Similarly, it might be argued that a person avowing her mental states is epistemically entitled to her pronouncements inasmuch as her avowal is made

“in accord with norms of reasons,” provided only that in avowing, the subject is being *reasonable*, as opposed to being prepared to offer reasons.

We now need to consider what would give one a special entitlement to her avowed self-ascriptions. Consider self-ascriptions of thoughts and occurrent beliefs. There is an important difference between ascribing such states to oneself in an “alienated” or “theoretical” way and ascribing them to oneself from the “first-person perspective.” Ascribing to myself the belief that I am leery of my neighbor’s friendliness purely as a result of taking my therapist’s expert word for it is one thing. Ascribing that same belief to myself with the expressive directness of an avowal is another. To avow a thought is at least to endorse the thought, and to adopt an intellectual stand on the thought’s content. It is thereby to incur an intellectual commitment to the avowed thought equal only to the commitment one incurs when giving the thought a non-self-ascriptive expression. If I avow (in speech or in thought): “I believe the Democrats will win,” and I am apprised of facts that point against a Democratic victory, it becomes rationally incumbent upon me to alter my first-order belief. No such pressure arises in connection with ascriptions of beliefs to others or with “alienated” self-ascriptions. (I can easily ascribe to someone else a belief I think is false. And I can even do so in my own case, provided I adopt a theoretical or “third-person” stance.)

Avowing in general puts one in a special relationship to what is avowed. Avowing does not allow disowning the avowed state the way one can disown states ascribed to another, or even states ascribed to oneself in a detached, theoretical (or observational) manner. Although high-road proponents focus their treatment upon avowals of propositional attitudes, the basic idea seems applicable also to avowals of non-doxastic states—occurrent wants, preferences, emotions, feelings, and even sensations.³³ Just as an avowal of a belief or thought involves taking a direct intellectual stand on the subject matter of the thought, so an avowal of an occurrent desire, emotion,

or feeling, involves genuinely claiming one's state. And it can be argued that engaging in rational deliberation or critical reasoning requires being able to ascribe to oneself not only beliefs, thoughts, judgments, etc., but also emotions, feelings, and sensations from the "first-person" perspective – or "in the avowing mode," as we prefer to say.

A parallel point can be made regarding practical deliberation.³⁴ To be a practical deliberator, it is obviously not enough that one *have* first-order desires, intentions, feelings, etc. One must also be able to recognize that one has them. One has to be able to self-ascribe those states as part of one's deliberative processes. Arguably, it is not enough that one be able to self-ascribe those states in a theoretical or observational way, as one does in the case of ascriptions to others. It seems that at least some of the self-ascriptions that serve as steps in practical deliberation must have the expressive force of avowals. The motivational role played by these self-ascriptions may depend at least in part on their being issued as avowals, rather than as mere reports on one's present mental states.³⁵

If it were true that being a rational and practical deliberator required being in a position to self-ascribe mental states in the avowing mode, then the special entitlement that grounds privileged self-knowledge could be seen as follows. As an avowing subject, I would be entitled to the self-judgments that my avowals semantically express, insofar as these self-judgments are necessary elements in my rational and practical deliberation. This entitlement would be derived not from contingent facts about the reliability of my pronouncements as indications of my present states (*cf.* the low-road above), but rather from the assumed fact that I am a rational and practical deliberator, plus the apriori derived fact that the truth of the self-judgments semantically expressed by my avowals is necessary for rational and practical deliberation. The derivation could take the following, "transcendental" form:

If rational and practical deliberation is possible, deliberators' own avowals of present beliefs, thoughts, preferences, hopes, feelings, etc. must enjoy a special status. These avowals enjoy a special status only if the deliberator possesses a special entitlement to the judgments expressed by those self-ascriptions. Since rational and practical deliberation is possible, we must possess a special entitlement to the judgments semantically expressed by our avowals.

There is much to be said in favor of the high-road approach. It clearly speaks to the intuition that in some sense avowable self-knowledge is not optional, and that avowals hold a special place among the things we can be said to know. Authors that we consider to be representatives of the high-road approach point out connections between avowals and rational deliberation, action, free agency, and so on, that seem both interesting and important to understanding the role played by avowable self-knowledge in our epistemic lives. This may well represent an improvement on the stark low-road approach described earlier. Were we to take issue with the high-road approach, this would be because we shy away from some of the stronger claims these authors make in the course of fleshing out these connections. As regards the Burge-inspired proposal considered here, we suspect that, if avowals are apt to serve as crucial steps in rational and practical deliberations, this may be *because* they represent things we are in a special position to know, rather than the other way around. Furthermore, some proponents of the high-road approach connect their apriori reasons for thinking that we must possess self-knowledge with the idea that there is a “constitutive” relation between being in mental states and making true avowals (so if S sincerely avows being in a mental state M, then S must be in M, and vice versa, at least when certain general conditions are assumed to hold).³⁶ But we think that, even if the possibility of an individual who is “chronically unreliable” in her avowals is somehow ruled out on more than purely empirical grounds, there can be no apriori guarantee for the truth of any particular avowal.³⁷ But we will not now attempt to develop further such objections to the high-road approach.

5.3 A “Middle Road” to Self-Knowledge

It is interesting to note that both the low-road and the high-road approaches ground the epistemic warrant of avowing subjects in something that is external to the subject's epistemic perspective in a particular situation. On the low-road approach, true avowals can amount to knowledge because of the *de facto* truth-conduciveness, which is due to their general reliability as indicators of the subject's mental states. On the high-road approach, true avowals count as knowledge courtesy of necessary connections between our capacity to avow and other central capacities we possess. On the particular high-road proposal we considered above, avowals constitute necessary steps in rational and practical deliberation. But it might be thought that, if I am said to have special knowledge of some of my states that no one else can have, it is partly because of a special relation that I have to what I am said to know, or due to my being in a special position to have that knowledge in the relevant situation. In consonance with the Neo Expressive account, both the low-road and the high-road approaches accept that this relationship need not involve my forming a self-judgment on some distinctive epistemic basis, or possessing reasons for my avowals, or otherwise being aware of what warrants me in making them. So we might wonder in what way my having knowledge of my present states of mind can still count as some kind of epistemic achievement on my part. One might argue that, even if my being warranted in issuing a self-ascription does not require me to have made any epistemic effort, my having self-knowledge must still be connected to something I myself have intentionally done in the particular situation. If so, it may be felt that the low-road approach and the high-road approach are both guilty, each in its own way, of taking self-knowledge out of the hands of knowing selves.³⁸

This gives us reason to try to articulate a middle-road approach, one which grounds the epistemic warrant enjoyed by avowing subjects in the special epistemic position they are in when avowing. In presenting the approach we will draw directly on

resources offered by the Neo Expressive account of avowals' security. Recall that on that account the special security of avowals derives from the fact that they are expressive of the avower's first-order mental states. When presenting the Neo Expressive account we appealed to a distinction between three senses of "expression": the causal, the action, and the semantic senses. Our main concern there was to separate off the semantic sense from the causal as well as action sense. This is because we wanted to make it clear that an avowal can ascribe a particular state to a subject, semantically speaking, without being the upshot of the avower's own higher-order judgment that she is in that state. In other words, an avowal can express the judgment that a subject is in some mental state without expressing the subject's own judgment to that effect in either the causal or the action senses. The low-road reliabilist approach seizes on the idea that an avowal does express the subject's mental state in the causal sense, and maintains that the causal reliability of avowals plus the fact that they express truth-assessable claims about the subject should suffice to give them the status of self-knowledge. By contrast, the middle-road approach calls upon the idea that an avowal expresses the subject's mental state in the action sense.

Whereas we can say of a subject's *pronouncements* that they are more or less reliable indicators of a condition she is in, properly speaking it is only the *subject* herself that we can describe as warranted in making those pronouncements, and thus credit with self-knowledge. As a first step toward "putting self-knowledge back in the hands of knowing selves," the middle-road proponent points out that avowing, properly understood, *is something the subject does*; it is an action—overt or covert, in public or in mind—that is performed by an agent. Beginning with the idea that avowing is a certain kind of an expressive action that a subject performs, the middle-road proponent tries to identify something in the character of this type of action that may yield a special epistemic warrant for avowing subjects.

In explaining avowals' special security, the Neo Expressive account appealed to the idea that avowals enjoy immunity to epistemic errors of identification as well as ascription. This immunity, we claimed, explains why we would normally not ask an avowing subject for her reasons, challenge her avowal, or offer a straightforward correction.³⁹ It should now be re-emphasized, however, that the immunity in question is not merely "grammatical." If avowals are immune to epistemic errors, this is not because of the semantics of the products of avowals (or lack thereof), but because of what is involved in producing them. The products of avowals are genuine self-ascriptions, which differ from grimaces and giggles in being semantically articulate and truth-assessable. A self-ascription is immune to epistemic error only if it is produced in the avowing or first person mode; and that has to do with what the *subject* is doing when avowing. What yields the immunity is the fact that the subject does *not* take epistemic routes to the ascribed mental state, but gives it direct voice.

Now, although we introduced the notion of immunity to epistemic error by way of proposing a non-epistemic account of avowals' security, we think it can be used to support the idea that subjects enjoy a special epistemic warrant when avowing. On the Neo Expressive account, avowing as such is a matter of expressing first-order mental conditions using expressive vehicles that ascribe the condition to the avower. From the subject's perspective, avowing involves conscious use of a semantically articulate self-ascription, in thought or in speech, so as to give voice to a condition she is in. Avowing is not merely something that happens to a subject, like a sneeze or the appearance of a rash. In the typical case, it is something a subject intentionally does—an act of *speaking* her mind. However, we maintain that when a subject is speaking her mind, she is in an epistemic position that is different from when a subject sets out to discover or ascertain the truth about this or that matter, including her present mental states. If I avow a mental state, in contrast to reporting on it, my present state is not an epistemic target for me, and

the self-ascription issued is not epistemically mediated. An avowing subject is not on a fact-finding mission. She simply gives voice to mental goings-on.

So, while avowing is something a subject does intentionally, there is no reason to suppose that in order to be able intentionally to express her first-order mental state, the avowing subject must form the judgment that she is in that state. To accomplish the expressive act involved in avowing, one must be a master of self-ascriptive expressive means, and use them intentionally so as to produce a self-ascription. But there is no reason to suppose that intentionally ascribing a state to oneself must be backed up by a prior epistemic act of recognizing or ascertaining the state one is in. As pointed out earlier, the semantic achievement of self-ascription need not be epistemically underwritten.⁴⁰ The self-ascriptions involved in avowing are produced in “epistemic innocence.” The middle-road proposal is that this epistemic innocence yields a certain kind of warrant for the subject who is speaking her mind.

To see the idea, consider again the phenomenon of immunity to error through misidentification (discussed above, Section 3). If I say or think, “I am sitting down” in the ordinary way, I am not in a position to wonder whether it is I myself, or perhaps someone else, who is sitting down. This is because my bodily self-ascription is in no way based on an identification of the subject of the ascription. But now consider the question: “Do I know who it is who is sitting down?” Assuming it is true that I am sitting down, am I epistemically warranted in accepting that truth? The natural answer would seem to be “Yes,” even though I have engaged in no epistemic effort to ascertain the identity of the person who is sitting down. Of course, I may be wrong—say, if unbeknownst to me, my brain was hooked up to someone else’s body, and he was the person sitting down. In that case, I could not be said to know that it is I who was sitting down—simply because it would not be true. But the mere possibility of such a case is not enough to defeat my

warrant in the routine instance, nor does it trump my claim to *know* that it is I who is sitting down, when in fact it is.

If I am warranted in holding true that it is I who is sitting down, this is not because I have ascertained that the person sitting down is I. My epistemic situation is not one in which alternatives that could defeat my claim to knowledge come into play. From my perspective, there are no alternative candidates (*viz.*, for who it is who is sitting down) among which I must be able to make a reasoned choice, if I am to be credited with the relevant knowledge. My warrant in such a case comes from the fact that I cannot *go* wrong, epistemically speaking, even though, of course, I can *be* wrong (*i.e.*, I can hold-true that it is I who is sitting down when it is not). I am, we might say, *warranted by default*.

The middle-road approach proposes that, inasmuch as avowing subjects are immune to errors of ascription, as well as misidentification, they are warranted by default in issuing mental self-ascriptions. Questions from the perspective of an avowing subject about whether he is feeling happy rather than tired or is scared of the dog rather than the cat or is thinking that *q* rather than that *p*, simply do not arise. They *could* arise, if he were trying to figure out the truth about what is going on with him, in which case alternatives would become epistemically relevant. But insofar as he issues the self-ascription in the course of avowing—in simply speaking his mind—alternatives do not come into play. This is not to say that the self-ascriptions *must* be true; avowals can issue in false self-ascriptions. It is possible to avow having a sensation or feeling or belief that *p* when one in fact does not.⁴¹ Thus, the middle-road proposal is not that avowals are *true* by default. So it should not be confused with the Default View, which builds the correctness of avowals into the truth-conditions of mentalistic ascriptions through a default assumption (see above, sec. 2). Rather, the proposal is that, given the epistemic position an avowing subject is in, he is warranted in holding true what his avowal says,

despite the fact that he has not made an effort to rule out the possibility that what he says or thinks is false. The fact that what he says or thinks about his condition may be false does not defeat the subject's warrant for the self-ascription. Avowals' truth is not logically guaranteed; but that does not mean that to have knowledge of what they say subjects must actively do something epistemic to ensure they "got it right."⁴²

Avowable self-knowledge is not the only case where the idea of default warrant can get purchase. Consider a different (though related) case: that of knowing what one is doing in the course of doing it. You ask me, "What are you doing?" I say, "I am beating an egg", "I am drawing a house," "I am planting a bulb". As I give these answers, it is plausible to credit me with holding true the relevant claims, though in this type of case, too, it would seem odd to suppose that I am of the relevant *opinions*. And it would also seem out of place to ask me *how* I know what I am doing, or expect me to have reasons for the beliefs, or be able to justify them. Yet we normally credit people with knowledge of what it is they are doing, in the course of doing it. Note that such knowledge is by no means infallible. There may not be eggs in the bowl; what comes out on the paper may be a tower, not a house; what I am putting in the ground may be a pebble, not a bulb. (And, of course, I may be wrong in even more radical ways.) So in each such case, you may not credit me with knowledge of what I am doing, but that is because of the *falsity* of my claim, *not* because I am not warranted in making it. Falsity always trumps a knowledge claim, since one cannot know what is false (hence the "truth" condition on knowledge). But what is at issue is whether one can know that *p*, when *p* is true, even if one has not ascertained the truth of *p*, or has done nothing epistemic to rule out the possibility of *p*'s being false. And the case of knowledge of what one is doing may illustrate nicely that one can. If I say or think that I am raking the yard when I am in fact raking the yard, I do not merely happen to hold true what is true. Though I may not be able to produce reasons for my action self-ascription, it seems reasonable to suppose that

I am *warranted* in issuing it. But my warrant does not rest on any epistemic effort I have made to verify what I am doing. As the agent of my action, I am in a special position to pronounce on what I am doing, without having to engage in such an effort. My warrant is “executive” warrant, which is a kind of warrant by default.⁴³

In a similar vein, we may speak of “expressive” warrant as another kind of warrant by default. Expressing my first-order mental conditions is what I, as the *subject* of these conditions, am uniquely placed to do. Whereas others (as well as I) can ascribe mental states to me using various epistemic routes, no one but I can ascribe the relevant states to me in the avowing mode. Others can know of my state of mind only by forming a belief about it, and the belief will be warranted only if it rests on some epistemic basis. But in my own case, I can give voice to my state using a semantically articulate self-ascription. And when I do so, I am immune to the kinds of epistemic errors that could otherwise defeat my knowledge of what state I am in and am thus warranted by default in holding true that I am in that state. Only I am in this kind of position—to have knowledge of my current states of mind through speaking my mind.

Of the three approaches to self-knowledge discussed above, the middle-road approach seems to be the closest in spirit to the Neo Expressive account of avowals’ security. It goes beyond the mere contingent reliability of the products of avowals as indicators of mental states, but it remains shy of claims about the apriori necessity of avowable self-knowledge. It connects the possibility of possessing privileged self-knowledge with the special character of acts of speaking one’s mind. However, at least as presented so far, the middle-road proposal underplays important features of avowals highlighted by the other approaches: the truth-conduciveness of avowals, on the one hand, and the connections between avowals and central human capacities, on the other hand. The first feature seems crucial to the epistemic credentials of avowable self-knowledge (so described); the second seems crucial to its privileged status. If we are

seeking a complete answer to our question (ii)—viz., “What qualifies avowals as articles of *knowledge* at all, and what is the source of the *privileged status* of this knowledge?”—we might in the end need to draw on ideas taken from all three approaches. In particular, a final account might benefit from further study of the similarities and differences between avowable self-knowledge and other cases types of knowledge that do not appear to fit the straightjacket of the traditional JTB analysis, such as perceptual knowledge and knowledge of one’s actions, as well as knowledge through memory and testimony, and apriori knowledge. Developing such an account would require a fuller epistemological investigation than we have space for here. Nonetheless, we would like to conclude by outlining the shape such an answer might take.

On the Neo Expressive account, avowals’ security is not due to subjects’ possession of a special way of knowing; it is *not* a matter of their deploying special epistemic methods. Rather, it is connected to being a subject who *has* mental states and who is a competent user of self-ascriptive vehicles for the expression of those mental states. A subject who avows being happy to see her friend, like a subject who gives a hug, gives expression to her joy in the action sense. But the avowal, unlike the hug, is a semantically articulate self-ascription; it tells of a condition the self-ascriber is in by semantically expressing that condition. Still, the self-ascriptions produced in avowing come directly from the subject’s present condition, unmediated by processes of belief formation. So, as emphasized by the low-road approach, when the self-ascriptions are true, they do not merely *happen* to be true—they are highly reliably correlated with being in the ascribed mental state. Furthermore, in contrast to other epistemically unmediated pronouncements, such as proprioceptive reports or self-ascriptions of one’s present actions, avowals provide not only a reliable but also a crucial source of truth. Others’ epistemic access to a subject’s mental state normally depends on the subject’s avowals. I do not usually need you to tell me whether you are sitting or not, or where your arms are;

and at least in basic cases I can judge for myself what it is you are doing by just looking. But, unless you speak your mind, even some of your most basic mental states may elude me.

On the other hand, the avowing subject herself need not rely on her own avowals in order to have knowledge of her present states of mind. In avowing, she knows the state of mind she is in. As the subject is avowing (in speech or silently), she holds true something that may well *be* true, and, moreover, when it is, it does not merely happen to be true. Furthermore, as emphasized by the middle-road approach, when speaking her mind the subject is acting in “epistemic innocence,” and is thus exempt from the need to warrant what she says by ruling out the possibility that her avowal is false. Given the expressive nature of her act, the subject’s avowal is not only likely to be true, it enjoys warrant-by-default.

Elsewhere, we suggested that the security of avowals might be associated with a special *first-person privilege*.⁴⁴ First-person privilege is something each of us enjoys specifically with respect to the mental states that one undergoes at a given moment. It is exercised when one is speaking one’s mind. There *is* something special about such exercise, something that the high-road approach tries to capture. When I speak my mind, *I proclaim the very states—the thoughts, hopes, wants, pains, etc.—that move me in thinking and in acting, at the same time as I ascribe those states to myself*. My proclamations are not themselves backed up by reasons; but they offer up the very states that can *provide reasons* for what I think and do. I lay bare the inner springs of my action, so to speak. In this respect, avowable self-knowledge is different from other kinds of knowledge, including non-observational knowledge of one’s bodily states, as well as immediate knowledge of what one is doing. Inasmuch as avowing involves the exercise of their first-person privilege, avowing subjects may be described as genuinely

knowing selves, and their avowals represent what deserves to be called privileged self-knowledge.⁴⁵

ENDNOTES

REFERENCES

- Anscombe, G. E. M. 1975. "The First Person." In Samuel Guttenplan, ed., *Mind and Language: Wolfson College Lectures, 1974*. Oxford: Oxford University Press: 45-65. Reprinted in Cassam 1994: 140-59.
- Armstrong, David. 1968. *A Materialist Theory of Mind*. London: Routledge.
- Bar-On, Dorit. *Speaking My Mind: On the 'Grammar' and Epistemology of "I"-Talk*. Unpublished manuscript.
- Bar-On, Dorit, and Douglas Long. 2001. "Avowals and First-Person Knowledge." *Philosophy and Phenomenological Research* 62: 311-35.
- Bilgrami, Akeel. 1998. "Resentment and Self-Knowledge." In Crispin Wright, Barry Smith, and Cynthia MacDonald, eds., *Knowing our Own Minds*. Oxford: Oxford University Press: 207-41.
- Boghossian, Paul. 1989. "Content and Self-Knowledge." *Philosophical Topics* 17: 5-26.
- Burge, Tyler. 1994. "Individualism and Self-Knowledge," in Quassim Cassam, editor, *Self-Knowledge*. Oxford: Oxford University Press. 1994: 65-79.
- Burge, Tyler. 1996. "Our Entitlement to Self-Knowledge." *Proceedings of the Aristotelian Society* 96: 1-26.
- Burge, Tyler. 1998. "Reason and the First Person." In Crispin Wright, Barry Smith, and Cynthia MacDonald, eds., *Knowing our Own Minds*. Oxford: Oxford University Press: 243-70.
- Bratman, Michael. 1997. "Two Faces of Intention." In Alfred Mele, ed., *The Philosophy of Action*. Oxford: Oxford University Press: 178-203.
- Cassam Quassim, ed. 1994. *Self-Knowledge*. Oxford: Oxford University Press.
- Crimmins, Mark. 1992. "Tacitness and Virtual Beliefs." *Mind and Language* 7,3: 240-63.
- Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
- Fricker, Elizabeth. 1998. "Self-Knowledge: Special Access vs. Artefact of Grammar—A Dichotomy Rejected." In Crispin Wright, Barry Smith, and Cynthia MacDonald, eds., *Knowing our Own Minds*. Oxford: Oxford University Press: 155-206.
- Gettier, Edmund. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23: 121-23.
- Jacobsen, Rockney. 1996. "Wittgenstein on Self-Knowledge and Self-Expression." *Philosophical Quarterly* 46: 12-30.
- Long, Douglas C. 1994. "Why Machines Cannot Think and Feel." In Dale Jamieson, ed., *Language, Mind, and Art: Essays in Appreciation and Analysis, in Honor of Paul Ziff*. Dordrecht: Kluwer Academic Publishers: 101-19.
- Moran, Richard. "Introspection and Deflation." Unpublished ms.
- Moran, Richard. 1994. "Making Up Your Mind: Self-Interpretation and Self-Constitution." *Ratio* NS 1: 137-51.

- Moran, Richard. 1997. "Self-Knowledge: Discovery, Resolution, and Undoing." *European Journal of Philosophy* 5: 141-61.
- Moran, Richard. 1999. "The Authority of Self-Consciousness." *Philosophical Topics* 26:174-200.
- Moser, Paul. 1986. *Empirical Knowledge: Readings in Contemporary Epistemology*. Totowa: Roman and Littlefield.
- Putnam, Hilary. 1960. "Minds and Machines." In Sidney Hook, ed., *Dimensions of Mind*. New York: New York University.
- Pryor, James. 1999. "Immunity to Error through Misidentification." In *Philosophical Topics* 26, nos. 1 and 2: 271-304.
- Rosenthal, David, ed. 1991. *The Nature of Mind*. Oxford: Oxford University Press.
- Sellars, Wilfrid. 1974. "Language as Thought and as Communication." Reprinted in Wilfrid Sellars, *Essays in Philosophy and Its History*, Dordrecht, Holland: D. Reidel Publishing Co.: 93-117.
- Sellars, Wilfrid. 1988. "On Accepting First Principles." In James Tomberlin, ed., *Philosophical Perspectives*. Atascadero: Ridgeview: 301-14.
- Shoemaker, Sydney. 1968. "Self-Reference and Self-Awareness." In *Journal of Philosophy* 65: 555-67.
- Shoemaker, Sydney. 1996. *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.
- Strawson, Peter F. 1966. "Self, Mind and Body." *Common Factor* 4: 5-13. Reprinted in David Rosenthal, ed., *The Nature of Mind 1991*: 58-62.
- Smith, Barry. 1998. "On Knowing One's Own Language." In Crispin Wright, Barry Smith, Cynthia MacDonald, eds., *Knowing our Own Mind*. Oxford: Oxford University Press: 391-428.
- Vendler, Zeno. 1972. *Res Cogitans*. Ithaca: Cornell University Press.
- Wilson, Margaret. 1978. *Descartes*. London: Routledge & Kegan Paul.
- Williamson, Timothy. 1995. "Is Knowing a State of Mind?" *Mind* 104: 533-65.
- Wright, Crispin. 1998. "Self-Knowledge: The Wittgensteinian Legacy." In Crispin Wright, Barry C. Smith, and Cynthia Macdonald, eds., *Knowing Our Own Mind*. Oxford: Oxford University Press: 13-45.
- Wittgenstein, Ludwig. 1958. *The Blue and Brown Books*. Oxford: Basil Blackwell.

¹ In ordinary parlance, "avowal" is reserved for utterances made in speech. But it should be granted that there are analogues of such utterances made to oneself outside communicative contexts. We shall often speak neutrally of self-ascriptions, as opposed to self-ascriptive utterances or thoughts. In keeping with common practice, we shall use quotation marks to cite the content of ascriptions, without presupposing that ascriptions made in thought must consist in one's talking to oneself in a language.

² It is actually not clear that the Cartesian account can preserve Epistemic Asymmetry without violating Semantic Continuity. Very briefly, the Cartesian account requires dividing ordinary ascriptions, whether in the first- or third-person, into ones that refer to bodies and ones that refer to Egos. As pointed out long ago by Strawson (1966, in Rosenthal 1991, 59) numerous everyday ascriptions do not lend themselves to this sharp semantic division. (Consider, e.g. “John is writing a letter,” or “Mary is jumping with excitement”.) But, in addition, the Cartesian story would require considerable reinterpretation in order to preserve the validity of humdrum inferences such as: “If someone has had a lot of sleep, then he is not likely to be tired. D got a lot of sleep. So he shouldn’t feel tired.”

We are less concerned here with an accurate representation of the historical Descartes than with the so-called Cartesian dualist view. For relevant discussion of Descartes’s view see Wilson (1978), chap. VI.

³ Following Bar-On and Long (2001).

⁴ Following Bar-On (2001). The distinction between the two questions is discussed further in Bar-On (ms., chaps. I and VIII).

⁵ There is some affinity between our two questions and the two sub-tasks Elizabeth Fricker identifies for a philosophical account of self-knowledge (1998, sec. II). In this paper, we are concerned with possible ways of tackling what Fricker identifies as the second sub-task, which she largely sets aside.

⁶ This is true even of sophisticated incorrigibility views such as Shoemaker’s (1996, chap. 6). We do not defend our opposition to incorrigibility here. For some discussion and examples, see Bar-On (2001, sec. 6). (The issue is further discussed in Bar-On (ms., chap. VII).)

⁷ Of course, the materialist explanation claims only a contingent superiority for the “first-person” epistemic route. My access to my own present mental states is epistemically more secure because causally more direct. Though others’ access is normally less direct, it is not necessarily so. And it is also contingent that I have such a direct access to my (so-called) mental states. I could have similar access to some of my (other) bodily states. In fact, it may well be that I already have such access: through proprioception and kinesthetic sense. (See below.)

⁸ We offer reasons against the Materialist Introspectionist account, as well as other non-Cartesian, “Epistemic Basis” accounts, in Bar-On and Long (2001, sec. 2) and Bar-On (ms., chaps. III and IV). See also Shoemaker (1996, chaps. 10 and 11), Wright, Bilgrami, and Smith in Wright (1998) for related criticisms.

⁹ The Simple Expressive view is criticized by Wright (1998, sec. 9). The view is also discussed briefly in Bar-On and Long (2001, sec. 3), and in Bar-On (ms. chap. VI and appendix 3).

¹⁰ In offering the Default View, Wright is partially motivated by an analogy with other areas of discourse where it seems misguided to conceive of successful judgments as tracking a completely independent reality—e.g., color judgments, and, on some views, ethical judgments. In Bar-On (ms. chap. VIII), it is argued that the analogy fails to address the peculiarly “reflexive,” first-person character of avowals’ security. See also Moran (unpublished ms., esp. p. 20f.).

¹¹ For further discussion of the Default View, see Fricker (1998, 200ff.).

¹² The presentation of our preferred account in this section follows Bar-On (2001), which, in turn, is an attempt to motivate and develop what we call the “modified expressive” account in Bar-On and Long (2001, sec. 4). A full defense and development of the account is offered in Bar-On (ms.).

¹³ Shoemaker (1968) and Evans (1982). Both Evans and Shoemaker introduce their diagnosis by discussing a much-cited paragraph from Wittgenstein (1958, 66f.) in which he draws a distinction between the use of “I” as object and as subject.

¹⁴ For a very helpful recent discussion of the phenomenon of immunity to error through misidentification, see Pryor (1999).

¹⁵ That “I” refers is famously denied by Anscombe (Guttenplan 1975 or Cassam 1994).

¹⁶ For the distinction between two senses of identifying an object, see Evans (1982, chap. 7.2). Both Evans and Shoemaker regard demonstrative thought (as in, e.g., “That thing is moving very fast,”) as not requiring “thick” identification. In such a case, one can succeed in “latching onto” a particular object in virtue of bearing non-epistemic relations to the relevant target of reference, such as being able to locate it in space, or track it as it moves, and so on.

¹⁷ Indeed, it is sometimes argued that the immunity to error through misidentification of nonmental “I”-ascriptions must ultimately be inherited from the immunity to error through misidentification of some underlying avowal. If my nonmental self-ascription, “The wind is blowing my hair about,” is IETM, this is due to the fact that it rests on a mentalistic judgment regarding how, for example, my scalp and face feel, which is in turn IETM. For a recent source, see Wright (1998, 19f.)

¹⁸ By “brute epistemic error” we mean error that is due simply to the world failing to cooperate, rather than being due to some kind of failure on the subject’s part. The contrastive model here is of perceptual errors, in judging, e.g., that there is a cup in front of me, which can occur even when all my faculties are working perfectly and my use of language is in perfect order. Burge discusses this (1996, 13ff.).

¹⁹ On the self-verifying character of such thoughts see Burge (1994).

²⁰ For a congenial though different treatment of Moore’s paradox, see Moran (1994) and (1997).

On the present view, Moore’s paradox is represented as a special case of “expressive dissonance.” There are analogues for Moore-sentences involving non-doxastic states, e.g., “I like this painting a lot, but it’s horrible” or “I find this meeting exciting, but it’s very boring” or “I feel so hot” uttered or thought as one is shivering and rubbing one’s arms. We take the anomalies exhibited by such combinations to attest to the expressive character of the contained self-ascriptions.

²¹ The self-ascriptions issued in such acts can be said to express the relevant state in Sellars’ causal sense, inasmuch as they are direct causal upshots of the ascribed state, rather than involving the mediation of a higher-order belief. We should note that our “non-epistemic” understanding the notion of expressing in the action sense may diverge from Sellars’ own understanding of it.

²² We should emphasize that we are here appealing to the example of promising by way of analogy only. We do not mean to suggest that an avowal brings the relevant state of affairs into existence the way an act of promising is supposed to. The analogy, however, serves to call attention to a familiar and relevant distinction from speech-act theory between the act performed in an utterance (the “illocutionary act”) and its semantic content.

²³ In Bar-On and Long (2001, sec. 4.2) we refer to this feature of expressions as their “transparency-to-the-subject’s condition.”

²⁴ More generally, first-order expressions of mental states, whether natural or not, are not self-ascriptions; ipso facto they are not self-ascriptions that are apt to be true. For example, a sincere

statement that p normally serves to express one's belief that p, where the expression of the belief is not plausibly thought as "natural." But it does not ascribe the belief that p to the person making the statement. So, although the statement itself is truth-assessable, it is not a truth-assessable self-ascription.

²⁵ See, e.g. Gettier (1963) and Williamson (1995).

²⁶ Williamson considers the possibility of analyzing the concept of believing as a disjunction of the concept of knowing and the concept of "opining," where opining is incompatible with knowing (1995). Thus "x believes that P if and only if x either knows that P or opines that P. Our suggestion is that when x knows that P he believes it at least in the sense of holding it to be true. There is an even more minimal notion of belief reserved for so-called tacit beliefs. For a recent analysis, see Crimmins (1992).

²⁷ In fact, it might be argued that even in the case of perception, I need not always ascertain or discover that which I can properly be said to know. Suppose I am staring out of the window and a robin lands on a branch right in front of my eyes. Can I not be said to see the robin, and for that reason, to have perceptual knowledge that there is a robin in front of me, even though I have not, strictly speaking, made what would count as an observation that could justify me? To accept this is not to obliterate important contrasts between avowals and perceptual claims. It is still the case that if I actually say: "There's a robin on that tree," I would be straightforwardly open to correction and questioning by others, as well as doubt by me. Perceptual claims, unlike avowals, cannot be said to enjoy ascriptive immunity to error. (See above discussion in sec. 3.)

²⁸ The avowing subject would not have to know that his self-ascriptions tracked his mental states in order to be warranted in taking them to be true. Consistently with his overall "externalism," the present reliabilist could insist that the subject needs only to know what his mental states are—not know that he knows what they are. Therefore, he is required only to produce self-ascriptions that in fact reliably correlate with his mental states and does not have to establish that he does.

²⁹ For a vivid description of such a device see Putnam (1960, 144-46). The focus of Putnam's discussion is not on the question whether such a device would have knowledge but on the parallels between minds and certain types of machines.

³⁰ D. C. Long has offered an explanation for why it seems reasonable to hold that an inanimate device, however complex, is not a proper subject to which to ascribe mental states at all, let alone genuine knowledge of its own states (1994).

³¹ The proposal we articulate below draws mainly on ideas presented in Tyler Burge's influential work on self-knowledge. See especially Burge (1996) and (1998). We cannot here do justice to the richness and subtlety of his views; our concern is to articulate a non-deflationary view of self-knowledge that is compatible with the Neo Expressive account of avowals' security. For other apriori reasons for the necessity of self-knowledge that could feed into a high-road proposal see Shoemaker (1996, esp. chap. 2), Bilgrami (1998) and Moran (1999). Limitations of space prevent us from presenting these further reasons.

³² See also Sellars (1988, 301f.).

³³ Burge says: "I regard knowledge of one's sensations as requiring separate treatment from knowledge of one's thoughts and attitudes" (1996, 17 n.13). We think, however, that Burge's main idea can be extended, once one adopts our Neo Expressive account of avowals.

³⁴ Though Burge speaks explicitly only of rational deliberation, the extension to practical deliberation that follows seems entirely in the spirit of his discussion.

³⁵ It may be for this reason that the job of a therapist who tried to persuade a patient that she hates her brother would not be regarded as complete until the patient actually came to avow the emotion. (See Moran (1997, sec. IV.) For the general contrast between theoretical knowledge of one's attitudes and the kind of knowledge reflected in avowals see also Burge (1996, 21ff.).

³⁶ Thus, Bilgrami tries to establish that the following condition holds "under the condition of responsible agency": If S believes/intends/desires/... that p, then S believes that S believes/intends/desires... that p, and vice versa. (1998, 211, 219, 222). It is not clear to us whether Burge would subscribe to such a "constitutive" thesis. This will partly depend on whether avowals can serve the role he assigns them in rational deliberation without being true.

³⁷ That particular avowals (phenomenal, as well as intentional) can be false seems obvious. We consider an example in Bar-On and Long (2001, 334). Bar-On (ms. chap. VII) examines a variety of "expressive failures" that issue in false avowals. Rejection of all constitutive theses of the above form marks one important way in which our Neo Expressive account departs from other recent expressive accounts. (See e.g. Jacobsen 1996.)

³⁸ The present point shares the flavor of some objections against epistemological externalism. But we are not here motivated by wholesale opposition to that approach, much less by wholesale endorsement of epistemological internalism. The worry raised here may have no force against those who would insist that the avowing subject has a reasoned belief about herself (belief in the opining sense) after all. But such insistence is not compatible with the Neo Expressive account of avowals.

³⁹ Another part of Epistemic Asymmetry is the strong presumption that avowals are true, which we have also explained by appeal to the expressive character of avowals—the fact that the self-ascriptions issued are taken to come directly from the subject's present mental condition. It is this aspect of the Neo Expressive account that the reliabilist focuses on.

⁴⁰ What constitutes the relevant act's being intentional in such a case is a legitimate question that would require an excursion into the philosophy of action, which we cannot undertake here.

⁴¹ On the Neo Expressive view, this would be due to expressive failures, which are to be explained psychologically rather than epistemically. (See footnote #37 for references.) Note that the present proposal requires denying that the mere fact that a subject issues a semantically articulate self-ascription means that he has identified in herself the relevant state. As argued in Section 3, semantic achievement need not be epistemically underwritten. Just as you can refer to yourself without any act of identifying yourself, so you can make a genuine (true or false) ascription of a mental state to yourself without any act of ascertaining the presence of the state in yourself.

⁴² Note that this does not mean the middle-road view collapses into the view that true belief suffices for knowledge. Avowals, on this view, do satisfy an additional condition: they are warranted (by default).

⁴³ Note that we are not claiming that self-ascriptions of action are like avowals in being immune to ascriptive errors. They are immune to error through misidentification, but, much like proprioceptive reports, they are open to doubt by the self-ascriber. As I say or think, "I am raking the leaves," I can perfectly sensibly wonder, "I'm doing something, but is it raking the leaves?" Note, too, that like proprioceptive reports, and unlike avowals, pronouncements on one's actions are straightforwardly open to challenge and correction by others. Such pronouncements rest in part on both proprioceptive and perceptual judgments, so they inherit the latter judgments' vulnerability to ascriptive error.

⁴⁴ In Bar-On and Long (2001) we replace talk of “first-person authority” with the concept of “first-person privilege.”

⁴⁵ Parts of this paper are based on a talk delivered by Dorit Bar-On to audiences in Chapel Hill, Tlaxcala (Mexico) and Virginia Commonwealth University. Thanks to members of these audiences, and especially to Bijan Parsia, Bill Lycan, Keith Simmons, and Gene Mills, for comments and suggestions that helped with the present version. We also wish to thank Rogers Albritton, Tyler Burge, and Richard Moran for earlier discussions concerning the epistemological issues that surround the Neo Expressive account of avowals.