



Bar-On on Self-Knowledge and Expression

Citation

Boyle, Matthew. 2010. Bar-On on self-knowledge and expression. *Acta Analytica* 25(1): 9-20.

Published Version

doi:10.1007/s12136-009-0075-z

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4879142>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Bar-On on Self-Knowledge and Expression

Matthew Boyle, Harvard University

1. Introduction

There are various standard ways of making our ability to know our own minds seem puzzling. One approach presents the difficulty as epistemological: how I can justifiably make various judgments about my own present mental state without looking for behavioral evidence, when I can make them about the mental states of others only on the basis of such evidence? Another approach invites us to puzzle about the deference accorded to certain sorts of assertions: why are my claims about my own present mental states treated as having some sort of special authority that other people's claims about those states are not taken to possess? Yet another focuses on the apparent inapplicability of certain kinds of questions to my claims about my present mental states: why is it, for instance, that if I say "I have a headache," it seems out of place to ask "How do you know?"? But there is also a way of presenting the problem of self-knowledge that focuses our attention, not merely on how we *know* about our own mental states or what we can *say* or *ask* concerning them, but on what those states themselves can *be*, given the special relation in which they stand to our awareness. In this version, the problem raised by our capacity for self-knowledge is not primarily epistemological or semantic, but metaphysical.

One way to bring out this metaphysical aspect of the problem is to observe that there is in the world a certain person about various aspects of whose condition I know "from the inside," where this means something like: *know simply by being* that person. This observation has two parts: first, the claim that I have substantive *knowledge* about this person's condition, and second, the claim that my being the one who is in the relevant condition *by itself* suffices for me to have this knowledge. Both the attractiveness of these two claims, and their strangeness, come out if we contrast some facts that seem to be available from this inward angle with some others that do not. Consider:

| | |
|------------------------------|--|
| knowing what I'm thinking | knowing what I'm wearing |
| knowing whether my arm hurts | knowing whether my arm is broken |
| knowing how warm I feel | knowing how much I weigh |
| knowing what I'm about to do | knowing where I'm about to land (if, e.g., I've just fallen off a roof) |

When I know the sorts of things mentioned in the left-hand column, what I know in each case is surely a real fact about a certain person, one that other people might know as well, one that can have genuine explanatory significance. Nevertheless, these facts contrast with sorts of facts mentioned in the right-hand column in a striking way: whereas the facts mentioned in the right-hand column are ones that can hold of me whether or not I know them to hold, the facts in the left-hand column seem to be ones whose holding is not independent of my being aware of their holding. It is not merely that I am in a specially good position to *tell* whether I am thinking about Amsterdam, or whether I feel a shooting pain in my arm; it is rather that the very idea of such facts obtaining wholly unbeknownst to me is hard to make sense of.

Mental states of these kinds seem to be conditions whose very obtaining involves their subject's knowing that they obtain.¹ But the idea of a condition whose obtaining involves its being known to obtain by its bearer is perplexing. If the fact that I bear a certain property is a substantive fact, one whose actually obtaining is distinct from its merely being thought to obtain, mustn't its obtaining be one thing and my knowing it another? And how can we

¹ Not everything we intuitively classify as a mental state has this sort of tight connection with self-knowledge. Nevertheless, for *certain* kinds of mental states, it is tempting to suppose that such a connection holds. Even this much is not uncontroversial, of course, but philosophers who would controvert it are already advocating a way out of the puzzle I want to describe. Before accepting their suggestion, it is worth giving our "Cartesian" intuitions a bit of scope here: only if we allow ourselves to consider them can we consider how much of a problem they actually pose.

Note that many of the standard counterexamples to mental "transparency" – unnoticed headaches, the perceptual states of long-distance truck-drivers as they drive "automatically" along a familiar route, etc. – are really counterexamples to the claim that present mental states must be the objects of conscious attention, not to the claim that they must be known. Quite generally: to know something is one thing; to be consciously attending to it is another. Note also that to hold that the obtaining of a condition *F* involves its subject's knowing it to obtain is not to deny that the subject may falsely believe himself to be in *F* when he is not: mental states may be *self-intimating* when they are present even though a person is not *infallible* in his judgments as to their presence.

understand my *justification* for taking myself to bear a certain property if my bearing it must involve my knowing myself to bear it? Indeed, if my bearing a certain property involves my knowing myself to bear it, what is it that I am supposed to know? The very *content* of such knowledge seems hard to explain when we begin to reflect on it.

Faced with such difficulties, there is pressure to give up one or the other of our two intuitive claims: either to suggest that the relevant condition's holding and my knowing it to hold are independent states of affairs; or else to deny that what is expressed in such self-ascriptions is substantive knowledge. Down the former path lie views which treat my knowing my own mental state as involving my being in a distinct, "higher-order" mental state, a state which presumably either results from my exercising some special epistemic faculty, or else is brought about by some reliable mechanism. Down the latter lie views which deny that self-ascriptions of present mental states express substantive knowledge, either because the existence of the relevant states is simply constituted by our willingness to self-ascribe them, or else because the apparent self-ascriptions do not really express judgments at all. To follow either path, however, seems to involve abandoning something intuitively attractive.

Dorit Bar-On's admirably lucid, meticulous, and suggestive investigation of our capacity for self-knowledge is, in effect, an attempt to find a more promising path through this terrain.² She frames the problem more cautiously than I have just done. Rather than supposing that self-ascriptions of present mental states normally express *knowledge* I have simply by being in the relevant states, she begins from the more neutral observation that such self-ascriptions, or at least a certain class of them, seem to possess a special "security": they are made without reference to evidence and yet are "issued with a very high degree of confidence and are not easily subjected to doubt" (p. 3). Following an established usage, Bar-On calls self-ascriptions belonging to this privileged class "avowals." Her question is how to account for their special security, and in particular, how to account for it without compromising their "semantic continuity" with third-personal ascriptions involving the same predicates.³ Her answer, in brief, is that we can understand both the special security of

² Dorit Bar-On, *Speaking My Mind: Expression and Self-Knowledge* (Oxford: Oxford University Press, 2004). Unless otherwise indicated, all page references are to this volume.

³ One question that might be raised at this point is whether Bar-On is right to assume that there must be a single

avowals and their semantic connection with other kinds of ascriptions if we understand them as claims that *express* the very states they ascribe – that is, as claims that stand to the conditions they ascribe in something like the relation in which moaning stands to being in pain, or spontaneously saying “He’s such a jerk!” stands (for a person who has learned this form of response) to feeling dislike for someone. Just as it is absurd to suggest that a subject who cries out in pain, or spontaneously calls someone a jerk, must have some ground for thinking that he is in a certain mental state, so, Bar-On suggests, we need not suppose that a subject who expresses such mental states in explicit self-ascriptions need have some distinct epistemic basis for doing so. For the point of her utterance might be “not to provide a descriptive report on [her mental state], but rather to share it, or air it, or give voice to it, or just ‘vent’ it” (p. 220), and thus might not stand in any greater need of an epistemic basis than do other expressive acts. We can accept this point, Bar-On suggests, without denying that avowals are truth-evaluable assertions that stand in all the usual logico-semantic relations to assertions made from other standpoints. For the claim that avowals are expressive pertains to the *act* of making an avowal, whereas the question of truth-evaluability concerns the *product* of avowing, “a linguistic (or language-like) token” (p. 251). Indeed, though expressive self-ascriptions are not made on some specific epistemic basis, they turn out, according to Bar-on, to meet plausible standards for counting as knowledgeable. So, contrary to what some earlier writers have supposed, the expressivist model need not rule out the ascription of knowledge to the avower.

This approach, if successful, would allow us to remain in the attractive middle ground, holding that I can have substantive knowledge of myself “no how,” simply by being in the relevant condition. It would allow us to avoid the twin temptations of (a) “the epistemic approach,” which represents self-knowledge as knowledge of an independent state of affairs supplied by some sort of special information-gathering faculty; and (b) the various approaches which deny that avowals express genuine knowledge, or that there is anything substantive here to know. I share Bar-On’s sense that this would be a desirable outcome. What I want to

account that covers all the kinds of avowals that exhibit some special security. I will not pursue this issue here, but for some grounds for doubting that we should expect a single, uniform account of all kinds of self-knowledge, see my “Two Kinds of Self-Knowledge,” *Philosophy and Phenomenological Research*, LXXVIII, 1 (2009), pp. 133-164.

do here is to raise some questions about her path to this result. Bar-On makes a forceful case for the thesis that the concept of expression must play a central role in a satisfactory account of self-knowledge, but I am not convinced that her own account sufficiently clarifies what this concept is and how it can play the required role. I will first explain what I take the required role to be (§2) and why I doubt that Bar-On successfully accounts for it (§3), and then raise a couple of questions about the notion of expression itself (§4). Along the way, I hope to give some grounds for thinking that my own less cautious formulation of the problem of self-knowledge has something to be said for it. For, I will suggest, Bar-On's ostensibly more neutral formulation fails to capture how the problem of self-knowledge confronts us when we approach the topic from the first-person standpoint. Looking at the problem from this standpoint will highlight its metaphysical aspect, and this in turn will help to bring out its importance: for if the problem of self-knowledge is not merely epistemological but metaphysical, solving it will require us, not merely to account for how mental states *are known* to their subject, but to rethink our conception of what sorts of things mental states *are*.

2. Avowals and their security

As Bar-On formulates the problem of self-knowledge, it is the problem of explaining the special security of avowals. To bring out my reasons for dissatisfaction with this way of framing the issue, it will be useful to give these two key notions more careful consideration. What exactly *are* "avowals," and what does their special "security" amount to?

Bar-On observes in her first chapter that the category of utterances that interests her is not specifiable simply by reference to features of surface grammar: there can be first-person present-tense ascriptions of mental states that possess no more authority or security than do mental state ascriptions made from other standpoints. Nevertheless, she suggests, there is "a familiar phenomenon, which can be characterized ostensively, as it were" (p. 26), one that comprehends self-ascriptions of both "phenomenal" and "intentional" states, and whose salient features include (a) that the relevant ascriptions appear not to be made on the basis of evidence, (b) that they are not subjected to certain kinds of ordinary epistemic assessment, and (c) that they are "issued with a very high degree of confidence and are not easily subjected to doubt" (p. 3). There are various ways in which these features of avowals might be thought puzzling and in need of explanation, but the puzzlement that Bar-On primarily

emphasizes concerns their special “security”: the special degree to which avowals seem protected from error and doubt. The remarkable thing about these utterances, she suggests, is that, “[w]hen compared to other non-apriori ascriptions, ... avowals are much more certain, much less subject to ordinary mistakes, significantly less open to a range of common doubts, and highly resistant to straightforward correction” (p. 10).

But is the striking feature of normal self-ascriptions of present mental states really their high *degree* of certainty or resistance to doubt and correction? Is it even true that our claims about our own present mental states are in general more *secure* than various other sorts of non-a priori knowledge we possess? I am quite certain that I am wearing pants, that I have two hands, that the sun is out, etc. The kinds of circumstances that would make me doubt these things seem at least as remote as those that might make me doubt my own avowals. For that matter, I am often quite confident about mental state ascriptions I make to other people on the basis of observation. If the reply is that I should not pick cases adventitiously but should contrast my general certainty in my avowals with my general certainty about other kinds of non-apriori assertions, then it seems that what I am being told to do presupposes that I can distinguish the relevant “kinds” of assertion by some other principle than my degree of certainty in them. And of course I can: I am familiar with the normal way of feeling entitled to self-ascribe a present mental state, and this way seems not exactly to possess a high degree of certainty or probable truth, but to be such that it is hard for me to make *sense* of certain kinds of doubt concerning it. This is what is brought out by the observation that it is normally out of place to ask me questions like “How do you know?”: there seems to be no “how” in the relevant cases. If confronted with the question how I know, I want to reply “Well, it’s *my* thought (my belief, my intention, my toothache, etc.)!” Indeed, the very language of certainty and doubt seems already to suggest the wrong kind of relation to such conditions. The dimension of my relation to my own beliefs or intentions is not certainty-or-uncertainty but something more like commitment-or-lack-of-commitment. And the dimension of my relation to my own pains is not certainty-or-uncertainty either: a toothache that I am not certain of feeling is either a very faint toothache or no toothache at all.

Now, in fairness to Bar-On, it should be said that she notes some of these very points. But in her subsequent discussion, these observations tend to recede into the background, while the focus remains on the fact that avowals are generally treated as having a high *degree* of

certainty and are not *often* questioned. The security of avowals, as she understands it, is primarily a matter of the avower's being *taken at his word* (not found mistaken, not subjected to doubt or correction), and her main question is why we treat avowals this way. And this way of framing the problem tends, in turn, to encourage us to adopt a certain point of view in thinking about the topic of avowals: it encourages us to imagine ourselves as *hearers* listening to a person making avowals, and to ask ourselves "Why should I take his word on the matter?" It encourages us, that is, to take up the topic of self-knowledge, not primarily from the standpoint of the subject who has the knowledge, but from the standpoint of another person considering why the utterances of that subject can be relied on.

What thus gets marginalized, I believe, is the question what it is like to *be* someone speaking his mind – not the one who takes the word, but the one who gives it. If we look at avowals from this angle, we are pointed, not primarily toward the question why we can *rely* on them, but rather toward the question how we can feel justified in *making* them as we do. What kind of condition could it be for which the question "How do you know?" would seem irrelevant *to the maker of the avowal*? The striking thing about avowals, after all, is not merely that we *do* sometimes make them nonobservationally. It is that we are perfectly confident of our *entitlement* to make them in this way, and would *not*, except in pathological cases, make them on the basis self-observation. What these observations suggest, I think, is that the remarkable thing about self-knowledge is not fundamentally its high degree of security but the *kind* of knowledge it is, a knowledge that stands to the condition known in a relation to which the categories of certainty and doubt are only marginally relevant. And this, I now want to argue, is an aspect of self-knowledge that Bar-On's account does not satisfactorily explain.

3. Expressivism and the first-person standpoint

So far I have simply been considering how Bar-On frames the problem of self-knowledge. Let me now turn to her proposed solution, the neo-expressivist account of the security of avowals. Bar-On notes that many earlier writers doubted whether an expressivist account of avowals can be reconciled with their semantic continuity with ascriptions of mental states made from other standpoints. According to Bar-On, however, this concern is not well-founded. Whatever the nature of the *act* of avowing, the *product* of this act is clearly a

linguistic token with a certain content. If the avower understands the semantic elements from which this token is constructed, and intentionally expresses himself in just these terms, then, Bar-On suggests, there is no reason to deny that he is making a contentful assertion about himself.⁴

I think that this is too quick, and that the earlier writers were right to be worried. To see the difficulty, we must note two points (each of which is granted by Bar-On). The first is a general point about linguistic understanding: namely, that understanding the content of an assertion requires grasping the systematic relationships between its truth-conditions and the truth-conditions of other assertions involving the same semantic elements. This is a way of expressing the requirement that Bar-On, following Gareth Evans, calls “the Generality Constraint,” and that she accepts as a requirement on any “genuine self-ascription of a contentful state” (p. 210). The thought is: You understand what you are saying in making a self-ascriptive utterance, and thus can be credited with having made a contentful assertion, only if you understand that you are referring to a person who can also be referred to in other ways, and saying something about him that can stand in logical relations to other kinds of claims. For instance, if I am to understand the content of the assertion “I have a toothache,” I must understand its connection with the thought “At least one person here has a toothache,” must understand what it would be for me to be identical to a certain objectively-described person, what it would be for some objectively-described person to have a toothache, etc. Now, add to this a second observation: that avowals are typically made *intentionally*. That is, these utterances are not just blurted out involuntarily, but are voluntarily made by speakers who know what they are doing. Again, this is a point Bar-On accepts.

If these two points are accepted, however, I think it is hard to see how an avower can

⁴ Although I will not pursue the point here, I should remark that this distinction seems to me more problematic than Bar-On suggests. I grant the intelligibility of the idea of a particular *act* of avowing, and the idea of a certain sentence-*type* that is “tokened” in such an act. But is there anything that is “the token” of that type other than the act of avowing itself? What is the token supposed to be? If it is *the utterance* of the sentence in question, that is surely the act itself and not something produced by the act (and something similar will apply if we allow talk of avowals “in thought”). But if the token isn’t the utterance, then what is it? I think these questions point to a problem for Bar-On’s strategy of isolating the expressive character of the avowing “act” from the semantic significance of the avowed “product.”

regard himself as *justified* in making avowals without self-observation. For if the avower is to count as making a contentful assertion, then he must understand what he is saying, and then our first point implies that he must understand that what he says implies certain propositions that can be justified by observation, and excludes others that can be so justified. He must, in short, understand that what he says involves commitments about what is observably the case. Furthermore, our second point implies that he does not just make the relevant self-ascription blindly or instinctively, but does so wittingly and willingly. But then how can he think himself justified in saying what he does, in spite of having no evidence concerning the observable state of affairs whose existence his utterance implies? Of course, he *could* make such assertions even if he did not think them justified: he could simply lie. But that is obviously not the normal case: normally, I have no doubts about my entitlement to say, without observation, that I have a toothache. The problem is to understand how this confidence can appear reasonable from the avower's own point of view.

The classical expressivist approach was to cut this knot rather than trying to untangle it. The idea was that a subject making an avowal need not be troubled by the question how his assertion is justified because he does not make a genuine, comprehending assertion; rather, he just unthinkingly manifests an acquired disposition to produce such-and-such noises when in such-and-such a condition. Bar-On rejects this response for failing to respect the fact that avowals are normally intentional assertions which are understood by their makers to stand, in semantical relations to other possible assertions. I think she is right to find this consequence unacceptable. But it is not clear to me how her "neo-expressivist" account solves the problem the classical expressivists were trying to avoid. She suggests that a neo-expressivist can avail herself of the notion of

an act, whether linguistic or non-linguistic, ... whose point is not to offer a descriptive report, make an assertion, or provide someone with information about the speaker's present thoughts, feelings, emotions or attitudes. Such acts can be performed with no particular purpose, and for no other reason than to put forth, or air, or give vent to one's present state. (p. 260)

But how should we understand the idea that an avowal simply "vents" the mental state it ascribes? If to say that the act of avowing is performed "to vent" the relevant state means that it is done because the agent knows herself to be in the relevant state and thinks this would be a

way of venting it, then this takes for granted the very thing we were aiming to understand: her ability to know that she is in the relevant state. But if this isn't what the phrase means, how should we understand it? If it means that she just blindly lets out these utterances when in the relevant condition, then we are back to the classical expressivist account. We need another alternative.

Once again, this problem stands out most sharply if we focus on how things look from the first-person standpoint, the perspective of the person making the avowal. On the classical expressivist view, although the avower himself utters various self-ascriptive sentences, his *cognitive* relation to these utterances is like the relation of a person on a beach to messages he finds in bottles that wash up on the shore: he may *understand* the content of the relevant communications, and may learn something from them, but his understanding is not involved in their *production*. This seems an unacceptably alienated picture of a person's relation to his own avowals: an avower must not merely understand his own utterances when he hears them produced; his understanding must be exercised in their production. But then the difficulty is not merely to account for how he can reliably come out with something true, but how *from his standpoint* the act of coming out with something can seem justified, given that he understands what he is saying and knows that he has no epistemic basis for the relevant claims. We need an account of what mental states *are* that explains how this can be reasonable. That is what I mean when I say that the problem of self-knowledge is fundamentally a metaphysical problem.

In an earlier exchange on this topic, Bar-On argued that this difficulty disappears if we take note of the distinction between expressing one's mental state in performing a certain intentional act and intentionally expressing one's mental state. Once this distinction is drawn, she suggested, we can admit that acts of avowal are *performed intentionally* without supposing that they are *performed with the intention of expressing one's mental state*; and then there is no tension between holding that avowals are made intentionally and recognizing that they are not made on the basis of any evidence about my mental condition:

When avowing ... we do not 'just blindly let out utterances when in the relevant conditions'. We engage in intentional behavior. *We* utter meaningful sentences; they don't just come out of our mouths... [B]ut although we engage in expressive behavior, and although this behavior consists in our doing

something intentionally, it doesn't follow that we're *intentionally* expressing anything.⁵

I think this distinction is a red herring. It is true that I can make a certain utterance intentionally, and thereby express something about my own mental state, although I do not make the utterance *with the intention* of expressing my own mental state. For instance, I can intentionally make a cutting remark about someone, and thereby express my envy towards him, although I have no intention of expressing my envy. But the problem I am raising does not depend on the assumption that, when one makes an avowal, one must do so with the intention of *expressing* one's present mental state. My question is this: When one makes an avowal such as "I feel a pain in my tooth" or "I am thinking of Amsterdam," does one, or does one not, intentionally *make a certain definite assertion about oneself*?

If it is said that one *does* intentionally make a certain definite assertion about oneself, then I ask: How can one regard oneself as justified in doing so? After all, if the avower understands what he is saying, then, as we have seen, he understands that his assertion implies the existence of a certain observable state of affairs, a state of affairs about whose existence he possesses, by hypothesis, no evidence. There is no need to assume that he intends to *express* his mental state in making this assertion; all that need be assumed is that he understands what he is saying, and intends to say something true about himself. If this is granted, then the question arises: How can he suppose that he is entitled to take the proposition he asserts to be true? And now it is not clear how appealing to the idea that his utterance "expresses" his mental state helps. If we suppose that the avower does not know that he is in a certain mental state which his utterance will express, then obviously he cannot draw on this fact in explaining to himself what justifies his assertion. But if we suppose that he does know this, then we have taken for granted what we were supposed to explain: his ability to know his own mental states.

On the other hand, if it is said that the avower *does not* intentionally make a certain definite assertion about himself, then I maintain that this leaves us with an unacceptably alienated picture of our relation to our own avowals, one that leaves the avower standing to

⁵ Quoted from Bar-On's "Reply to Matt Boyle and David Rosenthal," delivered at the meeting of the Eastern Division of the American Philosophical Association, Baltimore, MD, December 28, 2007, p. 5.

his own avowal in a relation essentially similar to the relation in which he stands to claims made by other people. I can see only two ways of denying that, when I avow “I feel a pain in my tooth” or “I am thinking of Amsterdam,” I intentionally make a certain definite assertion about myself: either I do not intentionally utter these words, or I do not do so with an understanding of what proposition my uttering them will express. But surely neither of these options can be what Bar-On intends. If I do not intentionally utter these words, then the fact that they come out of my mouth must, it seems, be a case of the operation of a blind disposition in which my understanding is not involved. This is unacceptably alienated: it leaves me in a position to know about my mental condition on the basis of *hearing* what I avow, but it does not represent my avowal itself as expressing knowledge. But if I do intentionally utter these words, but without understanding what proposition they express, then I do not comprehend what I am saying, and hence do not knowingly affirm what others would deny if they said of me “He doesn’t feel a pain in his tooth” or “He isn’t thinking of Amsterdam.” In short, semantic continuity is lost.

What I conclude from all this is *not* that we must make avowals on the basis of evidence about our own mental condition. I conclude that Bar-On has not yet explained how it can be reasonable for us to make avowals without evidence, as she rightly insists that we do. Explaining this will, I believe, require that we give a quite particular account of what sorts of things mental states *are*, and in this sense, it will require us to investigate the metaphysics of mental states. The idea that mental states are conditions that are *expressed* in our behavior may indeed be a crucial part of the explanation, but if so, I do not think Bar-On’s theory makes sufficiently clear how this idea helps to resolve the difficulty.

4. The notion of expression

The aim of the foregoing discussion has been, not to cast doubt on Bar-On’s claim that the notion of expression has an important role to play in accounting for the possibility of self-knowledge, but to suggest that there is more work to be done in clarifying this notion and the role it plays. In closing, I will say a few words about what I think needs clarifying here, and note a promising line of inquiry suggested by Bar-On’s discussion.

Expressivists about self-knowledge generally introduce the notion of expression by pointing to paradigm cases: groaning in pain, smiling with pleasure, blanching from fear,

bellowing with rage, etc. They then suggest that, despite the obvious difference between such “natural expressions” and verbal self-ascriptions, verbal self-ascriptions can come to play the same general *kind* of expressive role that such natural responses play. But what kind of role is this? What is the *principle* that governs our intuitions about what sorts of things count as examples of “expressive behavior,” and what exactly are we saying about behaviors when we characterize them as expressive?

One thing that is admirable about Bar-On’s book is that, rather than merely calling on the notion of expression to address other philosophical problems, as many earlier expressivists have done, she makes this notion itself an object of investigation.⁶ Her account of it includes several illuminating observations. For one thing, she clarifies and builds upon Wilfrid Sellars’s distinction between three senses of “expression”: an “action sense,” in which a person expresses her condition by intentionally doing something; a “causal sense,” in which a bit of behavior expresses a state by being “a culmination of a causal process beginning with that state”; and a “semantic sense,” in which a sentence expresses a proposition by being a conventional representation of it (p. 216). She also devotes a rich and interesting section of her book to speculating about the nature of “the commonsense notion of expression” (pp. 264–284). She maintains, however, that her account of the security of avowals is independent of the results of this investigation. For her official purposes, she suggests, it is possible to remain neutral on the analysis of the notion of expression, and indeed, to remain neutral on the question whether there can be a uniform analysis of this notion (pp. 245, 254, 261). All that is strictly necessary, she suggests, is that there be a similarity between the expressive role of verbal but non-self-ascriptive utterances such as “God, please no!” or “He’s such a jerk!” and that of self-ascriptive utterances such as “I hope that won’t happen” or “I hate him.”

To the extent that our grasp of the notion of expression remains dependent on our intuitive grasp of the principle implicit in a list of paradigm cases, however, it is not clear that we really understand what it is for an utterance to play an expressive role. We know something about this sort of role, namely that it is supposed to involve mental states in some sense “manifesting themselves” without the mediation of a recognitional judgment. But this

⁶ Related investigations are pursued in two other recent books: Mitchell Green, *Self-Expression* (Oxford: Oxford University Press, 2008), which Bar-On cites as an important influence on her thinking, and David H. Finkelstein, *Expression and the Inner* (Cambridge: Harvard University Press, 2003).

point about what expressive behaviors *do not* involve does not yet clarify what they *do* involve. And consequently, it leaves us without a clear understanding of how an intentional self-ascription *can* play an expressive role, given the difficulties noted in the previous section. Lacking a theoretical understanding of the notion of expression, we are not in a position to resolve the conceptual question how it is *possible* for comprehending, intentional, self-ascriptive assertions to function in this way. If this question can be answered, I suspect it must be through the sort of investigation that Bar-On places outside the bounds of her official account.

At an important moment in this extra-systematic investigation, Bar-On observes that there are many sorts of behavior correlated with particular mental states which we would not normally count as “expressive” of those states:

There are many different things a subject may do as a typical result of feeling tired, or hungry, or in pain, or anxious, or contented, or amused, and so on...
But we think of only a small subset of these behaviors as *expressing* mental conditions, in either the causal or the action sense. (Contrast, for example, someone taking an aspirin with her wincing or rubbing her temples.) (p. 269)

This seems both true and significant. Taking an aspirin may be a typical response to having a headache, but it does not *express* headache in the target sense, since it lacks the right sort of connection with having a headache: it is connected with having a headache only *mediately*, in virtue of the subject’s believing that doing this will help. And similarly, shivering is a kind of “behavior” typically associated with feeling cold, but it is not, I would think, an expression of feeling cold, since it does not seem to be in the right sense an *act* of the subject, and since it does not seem to have the characteristic *communicative role* of an expressive behavior. Another person may recognize that I feel cold by seeing me shiver, but getting people to recognize this does not seem to be the *function* of shivering in whatever sense it is the function of expressive behaviors to make a person’s condition manifest. These observations begin to bring out how our judgments about which behavior-types are expressive depend on our implicit understanding of relationships between the concept of expression and various other concepts – the concept of an *act* of the subject as distinct from mere reflex activity; the concept of an *unmediated* manifestation of a condition; the notion of the *function* of an act-type; and the idea of *communicating* that something is so as distinct from merely being a sign

that something is so. It should be clear that each of these concepts itself needs clarification: each has the feel of a notion that bears within itself a great raft of potential philosophy. My suspicion is that, if there can be a satisfying neo-expressivism about self-knowledge, it must come through a fuller consideration of the place of the concept of expression in this wider network of concepts.

I have focused here on doubts about Bar-On's version of neo-expressivism, but I am myself inclined to think that the notion of expression must play an important role in accounting for self-knowledge. For if a solution to the problem of self-knowledge requires an account of a kind of condition whose being involves its being known to its subject, it also requires an account of how the obtaining of such a condition can be a substantive fact about a subject, a fact knowable by others on the basis of observing what he does. Bar-On argues convincingly that the concept of expression must figure in such an account. A major part of the promise of her book, as I see it, lies in its bringing out the importance of this concept, and opening a path towards understanding it.⁷

⁷ For comments on an earlier draft, I am indebted to Rachel Cohen, Matthias Haase, Doug Lavin, Dick Moran, and Dorit Bar-On herself.