

Speaking My Mind: Expression and Self-Knowledge
(Oxford: Clarendon Press, 2004)
Reply to Matt Boyle and David Rosenthal
Dorit Bar-On

I'd like to begin by thanking the organizers of this session for their choice of commentators, and the commentators for the time and effort they've put into their comments. Given the time I have, I can't hope to address all the interesting comments they've made, so I'll confine myself to those that I take to be most accessible even to listeners who aren't familiar with the book.

Avowals' Security

The neo-expressivist view of avowals that I develop in the book can be thought of as a kind of 'middle position' between epistemic views and constitutivist views of so-called first-person authority. Like the constitutivist, I think that the security of our avowals is not simply a matter of our being contingently better placed to 'look in' on our mental states or our otherwise being able, as a matter of brute fact, reliably to pronounce on them. But, like defenders of non-Cartesian epistemic views, I deny that the security of our avowals is absolute, or a matter of conceptual necessity.

Both David and Matt have raised questions about the character of the security in question. In the book, I begin with a rough and ready characterization, which I gradually refine. I initially identify two components in avowals' security:

- (i) First, there's the thought that ordinary epistemic assessments (requests for reasons, correction, doubt, etc.) are out of place in the case of avowals. This contributes to striking *epistemic asymmetries* between avowals and other pronouncements.
- (ii) Second, there's the *quasi-a priori presumption* that spontaneous pronouncements we make about our present mental states are *true*. (I speak of a *presumption*, because like David, I think that it's defeasible; I speak of it being *quasi-a priori* because, like Matt, I think that it amounts to more than an empirical expectation that is responsive to contingent reliability.

The question I raise early on in the book is whether the commonsense view of avowals' security can be vindicated – whether and how this view can be *reasonable*. This requires offering an explanation of the alleged security. Since I also think (with both David and Matt) that our avowals often enough represent *self-knowledge*, I'm seeking an explanation of the security that is at least compatible with a non-deflationary view of basic self-knowledge.

The first component of security seems readily explainable by the traditional ('simple') expressivist view. If avowals are just glorified expressive moans and groans, then it makes sense that we wouldn't subject them to correction, doubt, requests for reasons, etc. But this expressivist view flies in the face of the *semantic continuity* of avowals with other pronouncements – for example, the fact that they are truth-evaluable (which is anyway required by the above presumption of *truth*). So the task for an explanation of avowals' security becomes more complicated: to explain epistemic asymmetry without compromising semantic continuity.

If the epistemic asymmetry in question was only the asymmetry between first- and third- person ascriptions, we *might* hope to be able to accommodate it at least to some extent by appeal to mechanisms dedicated to telling us about our *own* but not others' internal states. Indeed, we have a neat example of this sort of model. My self-ascriptions: "My legs are crossed" or "I'm standing up" contrast with my third-person ascriptions: "Jane's legs are crossed" or "John is standing up" precisely in that the latter are arrived at through external perception, whereas the former are typically (though not necessarily) arrived at through proprioception and kinesthesia. But avowals seem to me to be more secure than even such bodily self-reports. How to explain this greater security?

I start by appealing to the notion of immunity to error. As Shoemaker and Evans have noted, proprioceptive and kinesthetic self-reports share a certain epistemic feature with avowals. They are "immune to error through misidentification". Very briefly, if I say (or think) "My legs are crossed," for example, I can't sensibly wonder: "Someone's legs are crossed, but are they mine?" since I have no more reason for thinking that *someone's* legs are crossed than I have for thinking that *my* legs are crossed.¹

As a first step toward my neo-expressivist explanation of the security of avowals, I propose that we should understand avowals as immune not only to error through misidentification but also to *error through misascription*: When *avowing* "I am in mental state M," I have no more reason for believing that I am in *some* state or other than I have for believing that the state that I am in is mental state M. Of course, to say that avowals are immune to error through misascription is not to say that they are infallible or incorrigible. It's just to say that they are protected from a certain array of epistemic errors (and thus corrections) – a *much wider* array than other ascriptions, *including*, specifically, proprioceptive and kinesthetic self-reports (as well as mental self-ascriptions that are based on evidence or observation).

By itself, this more refined characterization does not yet give us a full explanation of their special security. We need to understand *why* avowals enjoy the additional immunity.² Moreover, since immunity to error, in general, is no guarantee of *truth*, we still need to explain the quasi a priori presumption of truth governing avowals.

Expressing, Reporting, and Moore's Paradox

It is at this point that the expressivist idea comes into my account. When I avow: "I'm thinking that there's water in the cup" I produce a certain token in the course of performing a certain *act* (which I can perform in speech or in thought). The *product* of my act of avowing in speech is a sentence token that *semantically expresses* (s-expresses for short) the proposition that DB is thinking that there's water in the cup. That proposition is distinct from the proposition that there's water in the cup, of course. For this reason, if I utter the Moore sentence "I'm thinking that there's water in the cup, but there's no water in the cup" the sentence I produce is not self-contradictory. However, for all that, when avowing "I'm thinking that there's water in the cup" I may be *expressing in the action sense* (a-expressing for short) the same mental *state* that I express when saying "There's water in the cup".

My distinction between act and product has a familiar precedent. When I say: "I promise to take you out to dinner tonight", the sentence I produce says that DB promises

to take the addressee to dinner on a certain night. But the act performed in issuing the sentence may not be that of *reporting* a promise, but rather an act of making one. Similarly, if I say “You will close that door”, I may not be *predicting* a future event but rather issuing a command (though the sentence I produce will still be true iff the relevant event takes place). I enlist this distinction, along with the distinction between s-expressing and a-expressing, to pull apart the *semantic* properties of avowals and their *epistemic* (as well as *action*) properties. Note that we already need to separate the semantic from the epistemic in order to understand the phenomenon of immunity to error through misidentification. If I were to say or think “My legs are crossed” based on looking in the mirror, though the self-ascription issued would be semantically the same as that issued through proprioception, my judgment would *not* be immune to error through misidentification. I claim, similarly, that understanding avowals’ special security avowals requires appealing to the kind of act performed when avowing, rather than focusing on the semantic content of avowals understood as products.

Going back to Moore’s paradox: I agree with David that Moore *sentences* are not self-contradictory; they do not s-express contradictions. Still, I maintain that if we regard a Moore sentence (or its analogue in thought) as *the product of an act of avowing*, we can offer the following diagnosis of its non-semantic anomaly. To avow “I’m thinking that there’s water in the cup” is to a-express one’s state of thinking that there’s water in the cup – the very state that one self-ascribes, and the same state that one could also a-express by simply saying “There’s water in the cup”. But going on to say under the same breath “There’s no water in the cup,” thereby a-expressing one’s state of thinking that there’s *no* water in the cup, would land one in an *expressive conflict*. There’s no logical contradiction here, because states of thinking are not the sort of things that stand in logical contradiction. But we can still preserve the sense of anomaly, which is what we want. Note that the neo-expressivist account predicts that similar anomalies abound. As David points out elsewhere, we have “Thank you but I feel no gratitude”; we can also have “[Yawn] but this is so interesting” or fanning oneself while saying “It’s not hot in here” (the last two of which involves no self-ascriptions).³

We now have on the table the key ingredients for my general explanation of avowals’ security. Portraying avowals as immune to errors of misascription gives us a handle on the epistemic simplicity and immediacy of avowals without having to compromise their semantic complexity. Avowals are not to be seen as upshots of recognitional identification of either oneself *or* the mental state one is in. Instead, we should think of avowing as ‘speaking from’ the mental state one is in. The speaking (or thinking) in the case of avowals is *self-ascriptive*; but this is just a feature of the expressive vehicle used in speaking your mind. Instead of saying “There’s water in the cup” you say “I’m thinking that there’s water in the cup”, instead of saying “That dog is scary” you say “I feel scared of that dog”, instead of saying “Ouch!” you say “I feel pain”, and instead of giving a hug you say “I’m so glad to see you”. Here we have the *performance equivalence* of which David speaks. But it seems to me that we need to go beyond merely acknowledging that there’s performance equivalence; we need to explain what underlies it, *especially* if our aim is to explain avowals’ security. The neo-expressivist account allows us to do so by providing not only for the epistemic simplicity and immediacy of avowals, but also for the quasi a priori presumption of their truth. For

if, when avowing, I a-express the very same mental state that my avowal attributes to me, then taking me to be avowing involves taking my avowal to be true.

Expression, Intentional Action and Reasons

Now both David and Matt had some misgivings about my use of Sellars notion of expression in the action sense (a-expression). On Sellars' characterization, "in the *action* sense, a *person* expresses a *state* of hers by intentionally doing something". To use a previous example, when I give you a hug, or say: "It's so great to see you," I express, in the action sense, my joy at seeing you. The notion of a-expression requires that a person do *something* intentionally. It does not require that what one does intentionally is *express*. One *can* intentionally express a mental state – for example, by *deciding* and *setting out* to give vent to a present emotion, instead of suppressing it. But the more basic case is one in which a person gives *spontaneous* expression to a present state of hers through performing some intentional act that doesn't have expression as its intentional aim or purpose.⁴ [And, I suspect that, on pain of regress, the more basic case still has to be one in which our *non-intentional* behavior causally expresses our mental states.] If so, then one could take issue with David's claim that, if I volunteer "I feel so cold", or "I hate this mess" (as I walk into my tween daughter's room), my reason for saying these things must be "the desire to let [someone] know, or even just express" that I feel cold or that I hate this mess, coupled with the belief that I can fulfill this desire in these ways.

Expressive acts, such as spontaneously smiling, stomping one's foot, making a rude gesture, giving a hug, saying: "What a mess," are typically not acts performed with a prior intention or with a specific goal or purpose in mind. Nonetheless, such acts seem to meet Anscombe's test for intentional action. They are voluntarily produced bits of behavior, where the person producing them *knows what she's doing* and where that knowledge is *nonobservational*. If our ordinary reason-giving practices are any indication, our reasons for such doings are typically *not* beliefs or thoughts we have *about* the expressed mental states but rather the states themselves: I'm giving a hug because I feel happy to see you; I'm sighing because I feel exasperated. But our avowals, I claim, can be performed for the very same sorts of reasons: when I avow "I am so happy to see you" I say this precisely because *I am so happy to see you*.⁵ Stomping my foot can be an intentional act of mine in which I a-express my anger, without the expressing itself being intentional; similarly, saying: "I'm mad at you" can be an intentional act of mine in which I (a-)express my anger, without my wanting or intending to express my anger through the utterance.

In the last part of his comments, Matt introduces the following worry:

If to say that the act is performed 'to vent' the relevant statement is to say 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. (p.6f.)

I think that what I said above about the intentional character of expressive acts provides precisely the alternative Matt demands. When avowing – as well as when expressing our states through non-self-ascriptive utterances, for that matter – 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.⁶ Our expressive *behavior* is intentional to the extent that we know what we’re doing (nonobservationally, if Anscombe is right); but 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.

I do hold that the self-ascriptive tokens we produce when avowing are typically produced intentionally, not just ‘blurted out’. And I agree with Matt that if I am intentionally to produce a self-ascription, I must understand that I’m “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” (p.6). But I deny that a meaningful production of a self-ascription must be epistemically underwritten by self-observation – that was my point in describing the immunity to error that our avowals enjoy.⁷

Matt goes on to object that I ignore a pressing question, namely: how, *from the standpoint of the avower*, the avowal can be reasonable even though it is made without any epistemic basis. (See p. 7).⁸ I think this runs together two issues that are worth separating. There’s the question what makes my *act of avowing* reasonable. When I speak of ‘rational cause’ and suggest that the expressed state is the rational cause in cases of avowing, *as well as* in cases of non-self-ascriptive expressive acts, I have in mind this question. Acts of avowing may be no more but also no less reasonable than other expressive acts. The fact that the avower has no *epistemic* basis for the avowal is not directly relevant to whether or not an *act* of avowing is reasonable. Now, when we avow, we produce an articulate self-ascription, a sentence that s-expresses a proposition, which can be true, believed and justified. So this raises an additional question: what, if anything, makes reasonable the *belief*, if any, that I may be expressing when avowing? The reasonableness in question *here* is that of *epistemic* justification. And the absence of epistemic basis does raise legitimate questions.

I argue, however, that these questions should be addressed separately, by a more directly epistemological investigation, and *after* we settle on the explanation of avowals’ security. I suspect that Matt is dissatisfied with my ‘divide and conquer’ approach because he thinks that there’s no explaining the security of avowals without explaining what makes us reasonable in the self-beliefs that they represent, and *that* in turn, he thinks, requires “an account of what mental states *are*”. This is what he means when he says that “the problem of self-knowledge is fundamentally a metaphysical problem” (p.7). But part of my aim has been to show that we *can* explain the security of avowals without committing at the outset to a specific metaphysical account of mental states, and *especially* not one that requires building self-knowledge or self-belief into their nature.

Self-Intimation, Avowals’s Security and Higher-Order Thought

It may seem natural to seek a metaphysical solution to the seemingly epistemological problem of self-knowledge if one’s starting point is Matt’s observation that “I have substantive *knowledge* about [the] condition of [one person in the world] and ... the

condition's obtaining *by* itself suffices for me to have this knowledge." (p.1 and compare p. 6). This putative explanandum is different from mine; it pertains to *self-intimation* rather than to avowals' security. With David (and others), however, I'm rather suspicious of self-intimation, and at any rate have sought an explanation of avowals' security that doesn't commit us to self intimation. The states we a-express with our avowals, I maintain, are not states that *necessarily* intimate their presence or character to us. On my view, one can be in a mental state without thinking that one is. Moreover, one can be in a mental state without expressing it.

It does seem to me reasonable, however, to suppose that the mental states we a-express are *conscious* mental states. So, as an alternative to my neo-expressivist account, David proposes that we preserve the idea that an avowal a-expresses *only one's higher-order thought* that one is in M – the very thought that, on David's HOT view, *renders M conscious*. The higher-order thoughts that render mental states conscious are, on David's view, thoughts that arise immediately, non-inferentially, and non-observationally as accompaniments of the first-order states. [And, if I understand David's view correctly, he thinks that, for various reasons, we shouldn't take these HOTs to be the upshots of special, reliable attention or belief forming mechanisms; he even denies that they need to be *causal* upshots of the first-order states.⁹] But this means that his proposal simply postpones, but does not answer, the questions surrounding avowals. If avowals only express 'unbidden' higher-order thoughts about our first-order mental states that simply come to our mind, how can it be at all reasonable to see them as secure pronouncements? And how can we take them to represent genuine self-knowledge?

Having said that, I agree with David that there are good reasons to suppose that avowals *do* a-express our HOTs, in addition to a-expressing the self-ascribed mental states. One of these reasons is connected to Matt's concern with self-knowledge: avowals, unlike other kinds of expressions, use *self-ascriptive* vehicles; this is what gets them caught up in issues of self-knowledge in the first place. But I remain convinced that it's the expressive tie to first-order states, not our higher-order thoughts about them, that holds the key to understanding avowals' security.

ENDNOTES

¹ For discussion of the phenomenon of immunity to error through misidentification, see Wittgenstein (1958b: 66–67), Evans (1982 (esp. ch. 7, sec. 2), Shoemaker (1968), and Wright (1998: 18–20). I discuss the phenomenon and its relevance to explaining the security of avowals in Ch. IV and VI of (2004).

² In the case of proprioceptive reports, the source of their immunity to error through misidentification has to do with our possessing special mechanisms for obtaining information concerning our *own* bodies. (See Evans (1982: Ch. 7).

³ David says: “if such sentences expressed incompatible thoughts, the sentences should carry some sense of contradiction” (p.5). If by “thoughts” we mean propositions (and by ‘express’ we mean s-express), then expressing incompatible thoughts will yield a contradiction (and not only in some sense). But if we mean “thinkings” – i.e., thought-*states*, which are what get a-expressed – then we don’t get a contradiction, since states do not stand in such logical relations, though we may have conflict in the domain of expressed mental states. (My explanation of what the expressive conflict on p. 218 of *SMM* is somewhat misleading in that it doesn’t emphasize that the (a-)expressed ‘beliefs’ that stand in conflict (but no contradiction) are belief-*states*.)

It’s worth noting that the neo-expressivist account puts us in a position to explain two other things that other accounts of Moore’s paradox don’t seem to explain. First, we can explain why “I’m thinking that there’s water in the cup” (which constitutes the first component of the above Moore sentence) is self-verifying despite not being necessary: the *act* of producing this self-ascription necessarily involves *a-expressing* the very state whose presence makes the self-ascription true, though the sentence produced doesn’t s-express a necessary truth. And second, suppose that you ascribe to yourself a certain belief solely on the basis of what your therapist has convinced you. You say: “I believe that my mother doesn’t like me”, though you still might be prepared to affirm: “My mother likes me”. Under those circumstances, it seems that you could sensibly say: “My mother likes me, but I *believe* she doesn’t like me”. This can be accommodated by pointing out that the second component of the self-ascription simply isn’t issued as an *avowal* but is rather issued as an alienated self-report.

⁴ Rosalind Hursthouse (19XX) gives various (often-cited) examples of expressive acts precisely to undermine the received dogma that we can only regard a piece of behavior as falling within the domain of the intentional if we can see it as backed up by a Davidsonian belief-desire pair. See also Green (2007).

⁵ Again, avowals are act with semantically complex, self-ascriptive products; but I don’t see why that necessitates regarding them as having belief-desire pairs as their reasons. This is, of course, not to deny that *sometimes* we produce self-ascriptions for more complex reasons. But we can also produce smiles and grunts for complex reasons.

⁶ But also: *we* give a hug, rub our eyes, yell. However, we don’t typically do these things on the strength of *believing* we are in certain state, *wanting* to express a state we are in, and *thinking* that producing these bits of behavior would be a good way to fulfill this desire. (And we can also engage in a ‘silent’ expression, where it seems even less plausible to suppose that our reason is a desire to let someone know, or even just to express ourselves.

⁷ Echoing Evans, I want to say: when I avow “I think it’s going to snow” or “I feel nauseous”, nothing more than the state of thinking, or feeling nauseous, is called for *on the epistemic side, or the side of awareness* for a subject to make an ascription of the state to herself. But certainly something more is called for *conceptually*. (This is precisely how Evans treats self-ascriptions of perceptual states such as “I see a tree”.) See Evans (1982: 231f.). And see *SMM*, p. 120 and p. 225.

⁸ Matt thinks that this is a consequence of the fact that I treat avowals “primarily from the standpoint of a *hearer*” (p.4). I actually think that both the notion of immunity to error and the notion of expression, as I deploy them in the book, are very well-suited to accommodating *both* the avower and the observer’s points of view, as I explain in various places in *SMM* (see esp. pp. XXX).

⁹ CITE