

# Oxford Handbooks Online

## Belief Self-Knowledge

Dorit Bar-On and Kate Nolfi

Subject: Philosophy, Epistemology Online Publication Date: Jun 2016

DOI: 10.1093/oxfordhb/9780199935314.013.51

### Abstract and Keywords

A fundamental puzzle about self-knowledge is this: spontaneous, unreflective self-attributions of beliefs and other mental states (*avowals*) appear to be at once *epistemically groundless* and epistemically privileged. On the one hand, it seems that avowals simply do not require justification or evidence. On the other hand, avowals seem to represent a substantive epistemic achievement. Several authors have tried to explain away avowals' groundlessness by appeal to the so-called *transparency* of present-tense self-attributions. After a critical discussion of two extant construals of *transparency*, this article presents an alternative reading of transparency (based on neo-expressivism about avowals) that explains, without explaining away, the apparent groundlessness of avowals. The article goes on to explore a way of coupling this alternative reading with a plausible account of how it is that ordinary avowals can represent genuine knowledge of present states of mind.

Keywords: Self-knowledge, beliefs, avowals, transparency, neo-expressivism, epistemically groundless, epistemically privileged

---

## 1. Introduction

We regularly attribute to ourselves states of mind, both in speech and in thought. Sometimes we do so spontaneously ("I feel so tired," "I wish I could go for a run right now," "I find this story so funny," "I hope he's not going to fall down," etc.). Other times, we do so in solicited responses to explicit questions about our current states of mind ("How do you feel?"—"Awful"; "Would you like to leave now?"—"Yes I would"; "What do you think of this play?"—"I think it's boring," and so on). Such self-ascriptions—*avowals*, as they are often called—have a distinctive epistemic profile. We ordinarily take others' avowals at face value. We strongly presume them to be true, we judge it to be inappropriate to challenge or correct them, and we regard our own avowals as a place of 'epistemic retreat': *qua self-attributions*, avowals are typically not to be questioned or

doubted, and we normally do not require that the avowing subject be able to supply justification for her self-ascriptive claims or that she be able to offer positive reasons on which these claims are based.<sup>1</sup> Yet we ordinarily also think that the avowing subject's self-ascriptive claims articulate the subject's knowledge of her own states of mind.

There is something puzzling here. On the one hand, it seems that our avowals simply do not rely on—nor do they require—positive justification. Their epistemic security does not seem to derive from an evidential basis; they are not—and need not be—underwritten by evidence or positive epistemic reasons at all. On the other hand, our avowals do seem to represent a kind of substantive epistemic achievement, and they seem to enjoy a certain *privileged* epistemic status: they appear to represent beliefs that are *especially apt to constitute genuine knowledge* of our own present states of mind. Thus, a fundamental puzzle about self-knowledge is this: our avowals appear to be at once *epistemically groundless* and also *epistemically privileged*.

Taken at face value, avowals' groundlessness can seem to invite *deflationism* about self-knowledge: the view that avowals are not rightly understood as genuinely knowledgeable claims or as expressing knowledgeably held beliefs about our states of mind. In an effort to avoid deflationism, some have recently tried to explain away avowals' appearance of groundlessness by appeal to a thought due originally to Evans and Moore (and later developed in a number of different ways by several authors).<sup>2</sup> The thought is that, when producing self-attributions of current states of mind, such as "I believe that it's raining," or "I'm having a visual sensation of something blue," we do not normally attend inwardly, as it were, to the contents of our mind (as introspectionist and acquaintance views would have it).<sup>3</sup> Instead, we attend to the same *outward* facts, objects, or properties that we would attend to if we were considering, for example, whether it's raining, or whether there's something blue in front of us. When making such self-attributions, we direct our attention at the world, not at the contents of our own minds; we look *through* our self-ascriptions to the worldly features at which they are directed. Call this feature of our ordinary avowals '*transparency-to-the-world*.'<sup>4</sup>

That avowals exhibit transparency-to-the-world, it is thought, helps explain why they *appear* to be groundless. For their transparency-to-the-world makes clear that, unlike other claims, avowals are indeed *not* epistemically grounded *in the facts that make them true* (viz. facts about the avowers' present states of mind). But this need not mean that they *are* epistemically *groundless*. As we'll see below, some authors who accept that avowals exhibit transparency-to-the-world suggest that, for precisely this reason, avowals should be seen as epistemically grounded *in the worldly features* (objects, properties, facts) that they are about.

Evans introduces the idea that avowals exhibit transparency-to-the-world (though not under that label) as part of his attempt to avoid the Cartesian conception of “self-knowledge as a form of perception—mysterious in being incapable of delivering inaccurate results” (1982: 225). In a much-cited passage, Evans observes that, if asked: “Do you think there is going to be a third world war?,” I must “attend to precisely the same outward phenomena as I would attend to if I were answering the question ‘*Will there be a third world war?*’.” He goes on to identify a “procedure for answering questions about what one believes” that can be encapsulated “in the following simple rule: whenever you are in a position to assert that *p*, you are *ipso facto* in a position to assert ‘I believe that *p*’” (*op. cit.*). And he adds: “the procedure only involves a direct consideration of the ascribed belief’s *content*” and the exercise of the same “normal abilities and dispositions for forming beliefs about the world” (*op. cit.*).

As we read Evans here, he takes it as given that ordinary self-ascriptions of beliefs (and other mental states) *are* instances of knowledge, and he appeals to their transparency to demystify the elevated epistemic status ordinarily assigned to them. His concern is to provide a non-Cartesian explanation of how conceptually articulate judgments concerning states of oneself can represent beliefs that are especially apt to constitute *knowledge about* those states, even though they are not epistemically grounded in peculiarly direct *consideration of* those states.<sup>5</sup> But Evans’ discussion leaves it a bit unclear how to construe his proposed ‘transparency procedure.’ In what follows, we canvass two recent construals of transparency-to-the-world—one *epistemic* and one *metaphysical*—both of which explain why it is that our ordinary avowals represent genuine knowledge of our states of mind by, in effect, simply denying that ordinary avowals are epistemically groundless. On the epistemic construal, what renders avowals knowledgeable has directly to do with their rational basis; on the metaphysical construal, it has to do with the nature of mental states. We will then present an alternative reading that both respects and makes sense of the apparent groundlessness of avowals. And, finally, we will explore one way of coupling this alternative reading with a plausible account of how it is that our ordinary avowals can represent genuine knowledge of our own states of mind.

## 2. The Transparency of Belief Avowals: Inferentialism vs. Reflectivism

In a series of articles, Alex Byrne has proposed that the best way to construe Evans’ thought that our avowals exhibit transparency-to-the-world is as pointing to a special but straightforward *epistemic procedure* via which we self-ascribe our own mental states. On

Byrne's view, one "*finds out* that one believes that it's raining by determining that it's raining: knowledge that one has this belief ... *rests on* perceptual *evidence* about the weather, not on perceptual evidence of one's behavior or anything mental. That is, one *reasons* from evidence that it's raining, to the conclusion that one believes that it's raining" (2005: 93, first three emphases added). Byrne's construal of Evans presumes that our self-ascriptions are epistemically privileged (i.e., especially likely to constitute knowledge) simply in virtue of their unique or distinctive provenance. Thus, for Byrne, uncovering why it is that our belief self-ascriptions constitute genuine knowledge involves identifying a unique and especially truth-conducive epistemic method via which we come to self-ascribe our occurrent beliefs (and other mental states).

The main idea of Byrne's view is as follows. In paradigmatic cases of belief self-attributions (as in Evans' "I believe there will be a third world war"), one arrives at the self-attribution via "an inference from world to mind: I infer that I believe that there will be a third world war from the single premise that there will be one" (2011: 203). In other words, one reasons in accordance with the doxastic schema:<sup>6</sup>

BEL: p

---

## I believe that p

A subject, S, reasons in accordance with the rule BEL when she believes that she believes that p *because* she recognizes that p obtains.<sup>7</sup> The 'because' here is the 'because' of the epistemic *basing relation*: the subject's recognition that p serves as both the causal and the rational ground for her *belief that she believes* that p.<sup>8</sup> But reasoning in accordance with BEL is unusual in that the content of the mental state that the subject ends up believing she is in as a result of reasoning in accordance with BEL is identical to the content of the rule's world-directed antecedent.<sup>9</sup> In this way, BEL seems to capture the transparency(-to-the-world) of belief self-attributions.

Byrne recognizes that the doxastic schema "is neither deductively valid nor inductively strong" (2011: 204). However, Byrne argues that the belief-forming method characterized by BEL is especially truth-conducive. This is because BEL is not merely reliable; it is *self-verifying*: if S believes that she believes that p as a result of following BEL, then her belief that she believes that p *must* be true.<sup>10</sup> And, moreover, beliefs produced in accordance with BEL are *safe*, in the sense that they could not easily have been false.<sup>11</sup>

Thus, on Byrne's view, it is features of the distinctive method that we employ in making second-order self-attributions of our first-order beliefs that explain why these self-attributions enjoy a kind of privileged epistemic status (in the sense of being especially apt to constitute knowledge).<sup>12</sup>

Byrne offers his inferentialist account as an alternative to accounts (most notably Richard Moran's and Matthew Boyle's) that use the transparency of belief to oppose 'detectivist' views of self-knowledge.<sup>13</sup> In a recent critique of Byrne's view, Matthew Boyle raises a powerful objection to Byrne's interpretation of transparency. To begin with, Boyle thinks that the conception of inference underwriting Byrne's discussion of BEL construes inference as nothing more than "a reliable process that deposits beliefs in [one's] mind" (2011: 231), and hence is severely and problematically impoverished. Crucially, for Boyle, we are equipped with, and regularly exercise a reflective capacity to know the way we go about forming our beliefs. Thus, Boyle writes, "I can *reflect* on why I draw a certain conclusion, and when I do, I can see (what looks to me to be) a reason for it" (ibid.). Boyle's point here is that when one considers the movements of one's own mind, one sees one's own cognitive transitions as movements from *reasons* for belief to belief itself. But, Boyle explains, it is "hard to see how the premise of Byrne's doxastic schema could supply me with a *reason* to draw its conclusion" (2011: 231, emphasis added). Given what we know about our own lack of omniscience, one cannot reasonably regard the mere fact that *p* as a reason for thinking that one believes that *p*.<sup>14</sup>

More importantly, however, Boyle thinks that the very idea that we need to rely on an *inference* from 'sheer propositions' about the world in order to know our present states of mind is wrongheaded (2011: 234). Just as wrongheaded, in fact, as the idea that coming to have second-order beliefs about one's first-order mental states should in the ordinary case require one to *observe* the presence and character of the latter. The inferentialist, Boyle thinks, presents a picture of basic self-knowledge that is just as "profoundly *alienated*" as the "spectatorial" picture offered by the introspectionist. For the inferentialist provides no principled reason for thinking that one has a kind of access to one's present states of mind that no one else can have.<sup>15</sup> Instead, Boyle suggests that, at least in the normal case, our knowledge of our own beliefs constitutes a kind of *engaged* awareness of our own minds: when all goes well, self-ascribing the belief that *p* *itself* involves reaffirming or reasserting one's first-order belief. And on Boyle's view, this is a feature of self-knowledge that any explanation of the privileged epistemic status of our beliefs about our own beliefs must respect.

As an alternative to Byrne's inferentialist reading of Evans' transparency, Boyle offers a *reflectivist* reading, which construes "doxastic transparency ... as a matter ... of shifting one's attention from the world with which one is engaged to one's engagement with it ...

an engagement of which one was already tacitly cognizant even when one's attention was 'directed outward'" (2011: 228). The reflectivist maintains that "in the normal and basic case, *believing P and knowing oneself to believe P are not two cognitive states; they are two aspects of one cognitive state*" (ibid., emphasis added). Far from representing the culmination of an inferential step from a 'sheer proposition' to a self-attribution, a subject's avowal "I believe P" represents "a coming to explicit acknowledgment of a condition of which one is already tacitly aware." Thus, "to pass from believing P to judging I believe P, all I need to do is reflect—i.e., attend to and articulate what I already know" (ibid.).

Boyle thinks that *no* account that appeals to an information-gathering faculty for producing beliefs about an independent state of affairs—whether inward *or* outward looking—could adequately capture the character and special epistemic status of self-knowledge. And he thinks that we can *only* avoid such appeal if we accept that the *facts* known when we know our minds are "ones whose holding is not independent of our being aware of their holding" (2010: 10).<sup>16</sup> What we need is "an account of what mental states *are* that explains how" it "can seem reasonable," from the standpoint of the avower, to make an avowal without any epistemic basis (2010: 16) and how an avower can "suppose that he is entitled to take the propositions he asserts to be true" (2010: 17). Boyle's account of self-knowledge is, then, intended to be metaphysical through and through: on his view, the privileged epistemic status of our *second*-order self-ascriptions is supposed to fall out of a complete account of the metaphysical nature of our *first*-order beliefs.

Although we agree with Boyle's objections to epistemic accounts, we have some general misgivings about his Cartesian-style appeal to the essential knowability of mental states—the idea that simply *being* in a mental state in some way brings in its train *knowledge* (if only tacit) that one is. But even setting aside these misgivings, we find it very unclear how building self-belief into the very nature of our first-order mental states could help satisfy Boyle's own demand for an explanation of what makes an avowal *reasonable* from the standpoint of the person who produces it. After all, the metaphysical nature of my first-order mental states, it seems, is something that need not be transparent to me. So even supposing it to be a metaphysical fact that, given that nature, I cannot *be* in a first-order mental state without being aware of it, if I really were to worry about what *justifies* me in believing that I believe that p, it's very unclear how that fact would provide me with the requisite *epistemic* assurance. At any rate, it would seem that whatever epistemic assurance I might derive from learning about the relevant facts about the metaphysical nature of my first-order belief would be no less alienated than that envisaged by an account like Byrne's.<sup>17</sup>

Boyle appears to avoid the risk of alienation by insisting that *all one needs* to do to move from believing that *p* to believing that one believes that *p* is “[*shift*] *one’s attention* from the world with which one is engaged to one’s engagement with it- ... an engagement of which one was already *tacitly cognizant* even when one’s attention was ‘directed outward’”; all I need to do “is *reflect*—i.e., *attend to and articulate* what I already know” (2011: 229, emphases added). But this means that, even accepting Boyle’s ‘dual aspect’ view of the metaphysical nature of mental states, the required epistemic justification of ordinary belief self-ascriptions does *not* simply fall out of this nature. For, on Boyle’s own account, it does not seem to be sufficient for one to *be* in the first-order state of believing that *p* in order to *know* that one is; some sort of attention shift and articulation is required to take one from mere tacit cognizance of one’s (doxastic) engagement with the world to belief self-knowledge. (At the very least, having belief self-knowledge requires an act of making explicit what is only tacit until one has engaged in the relevant ‘attention shift.’) And, pending some positive account of the cognitive character of this attention shift, it’s not yet clear how Boyle’s reflectivist account improves on, say, introspectionist or direct acquaintance accounts.

We thus believe that there are good reasons for seeking another alternative (non-introspectionist) account of what puts each of us in a special position to make knowledgeable pronouncements about our current states of mind. The desired alternative would attempt to capture the transparency of mentalistic self-ascriptions while avoiding both Byrne’s inferentialism and Boyle’s reflectivism.

### 3. Avowals' Security and 'First-Person Authority': A Neo-Expressivist Account

Both types of accounts of belief self-knowledge we have considered take the phenomenon of transparency to show that ordinary mentalistic self-beliefs enjoy a special form of epistemic support. We think this is due to the fact their proponents (Byrne and Boyle included) share a certain presupposition. This is the presupposition that the only way to account for the distinctive epistemic security of avowals (what sometimes comes under the umbrella of 'first-person authority') is to identify specific good-making epistemic features of the relevant self-ascriptions that either render avowals especially knowledgeable or confer a special kind of justification on the relevant self-beliefs. But this presupposition can (and, we think, ought to) be rejected. Following Bar-On (2004: 11ff.), we propose separating two questions:

- (i) What could vindicate the commonsense notion of first-person authority—or: how is avowals' distinctive epistemic security to be explained? And
- (ii) What renders the self-beliefs that avowals represent especially apt to constitute *knowledge*?

It has been traditionally assumed that the only way to vindicate the ordinary conception of first-person authority (thus answering (i)) is to identify specifically positive epistemic features that would reveal our mentalistic self-beliefs to be especially knowledgeable—that is, to answer (ii). This has led philosophers to reject offhand a certain type of *non-epistemic* explanation of first-person authority that is, nevertheless, compatible with a substantive, non-deflationist answer to (ii). Our view is that failing to recognize that the above two questions are distinct, and so perhaps most fruitfully answered independently, obscures the availability of an otherwise attractive account of self-knowledge. This account naturally comes into view once one accepts the *neo-expressivist* answer to (i) that we go on to summarize (which explains first-person authority *without* appealing to the positive features that render avowals instances of knowledge).

We can perhaps begin to free ourselves of the presupposition in question, and revise our understanding of the epistemic significance of the phenomenon of transparency, if we cease to focus excessively on self-attributions of belief as the paradigm case of transparent self-knowledge. We should recognize, first, that transparency-to-the-world is not exclusively characteristic of avowals of doxastic states, which (like beliefs) are responsive to reasons, open to reflective evaluation, and beholden to standards of rational justification. We can and do typically *also* tell whether we want or prefer x, are

annoyed at *y*, perceive *z*, plan to *q*, and even remember that *q*, by directly considering the intentional objects or contents of the relevant states. But, secondly, it's not the case that *all* avowals partake in the transparency-to-the-world of beliefs (whether understood Byrne's or Boyle's way). For an especially telling example, consider passing thoughts and unbidden desires. My authoritative response to the invitation "A penny for your thoughts"—for example, "Oh, I'm thinking about my grandmother"—could hardly be obtained through consideration of relevant worldly affairs. And it's implausible to think of a spontaneous, unsolicited avowal such as: "I'd love a cup of tea right now" as something *arrived at* through direct consideration of the world or reflective attendance to one's engagement with the world (even though it has overlapping content, as well as conditions of use, with a world-directed statement such as "A cup of tea would be nice right now").<sup>18</sup>

Excessive focus on avowals of beliefs and similarly reflective states may make it seem natural to take avowals' transparency to be a direct consequence of the fact that avowals *share* their epistemic grounds with the intentional states they self-ascribe. This seems much less plausible when it comes to avowals of states that are not themselves grounded in evidence or responsive to reasons. If I avow a desire for a cup of tea, there's a sense in which my focus is not on my desire, but rather on the cup of tea. But it does not seem plausible to suggest that I *arrive* at my self-attribution *via* consideration of the cup of tea and reflection on the reasons for favoring it. This is much less so when it comes to self-attribution of passing thoughts or sensations. And even when it comes to avowals that *are* transparent-to-the-world in the relevant sense, where we attend to the self-attributed state's intentional object (rather than attending to the state)—for example, attending to my neighbor's behavior to determine whether I'm annoyed at her—it's not obvious that the self-attribution rests on a rational evaluation of relevant features of the state's intentional object. To vary the example, attending to a dog in front of me, I may avow: "I'm scared of that dog"; but it does not seem plausible to suggest that my self-attribution is *based upon* my rational evaluation of the dog as scary.

Now, according to the *neo-expressivist* account (defended in Bar-On 2004)<sup>19</sup>, the distinctive epistemic security of avowals of beliefs and other states of mind is to be explained by appeal to the fact that avowals have the role of *expressing the self-attributed states*, rather than—or in addition to—expressing one's second-order *belief* about the presence of the state. An avowal doesn't serve to *report that* we are in a given state of mind; it directly expresses it. In other words, spontaneous, nonreflective, or unstudied self-attributions of present states of mind—for example, volunteering "I feel so tired!" or saying (or thinking) "I'd love some dessert" in response to a query—are importantly similar to *non-self-ascriptive* expressions of states of mind, whether verbal (e.g., saying, or thinking, "Dessert would be nice!") or nonverbal (e.g., yawning). As *acts*,

such avowals serve to directly reveal the very states of mind that the avowals understood as *products* (i.e., linguistic or mental representational tokens) ascribe to the avower.

Drawing on a distinction due to Sellars (1969), we can say that, for example, laughing at a joke, saying “This joke is so funny!,” or avowing: “I find this joke *hilarious*” are all performances in which an agent gives direct expression to a specific state of mind—expresses *in the action sense* (a-expresses, for short)—her amusement; though in each case she is using a different *expressive vehicle*.<sup>20</sup> A-expression is to be distinguished from expression *in the semantic sense*—s-expression, for short—which is a relation that holds between contentful tokens, such as sentences, and their semantic contents. (Some expressive vehicles—laughter, for example—do not s-express anything. And, as just noted, one can use sentences that s-express different propositions to a-express one and the same state of mind.<sup>21</sup>) When avowing, one performs a distinct type of act which serves directly to a-express *the very state that is self-scribed by the proposition that the sentence used s-expresses*. One is giving voice (out loud or silently) to the very same state of mind that is *named by* the expressive vehicle that one is using in her expressive act. Thus, like expressive acts more generally, avowals give direct voice to the avower’s states of mind and allow others to *see through* the avowals to the states of mind, though they do so using self-ascriptive vehicles. They are acts of *speaking one’s mind* self-ascriptively, *in lieu of* giving either non-linguistic or else non-self-ascriptive expression to one’s state of mind.

On the neo-expressivist account, an avowal such as “I’m scared of that dog” doesn’t (merely) *tell* of one’s fear the way a third-person report might; it directly *expresses* the self-ascribed fear.<sup>22</sup> Like other expressive acts that employ linguistic vehicles, avowals can rely on semantic and pragmatic features of the relevant vehicles to show the expressed states. In creatures like us, some of the communicative roles played by the more visceral showing and perceiving afforded by natural expressions, such as animals’ growls, bared teeth, grimaces, and so on, are taken up by spontaneous, competent use and immediate uptake of linguistic vehicles. (Swear words are one good example, but by no means the only one.) When it comes, specifically, to avowals, it might be argued that they *wear the states they are supposed to express on their linguistic sleeve*, as it were. An avowal such as “I wish we’d get some rain today” *explicitly names a kind of state of mind* (a hope) *and articulates its content (that it rain today)*, as well as *ascribing it to a certain individual*; it reveals the kind of state the avower expresses (as well as its intentional content, when it has one) through what the sentence expresses in the *semantic sense*.<sup>23</sup>

Insofar as all acts of avowing can be said to express—and thus to show—the ascribed states (in virtue of the self-ascriptive expressive vehicles they use), avowals can be said to enjoy a certain transparency—what can be described as their *transparency-to-the-*

*subject's-state*.<sup>24</sup> On the neo-expressivist account, *all* avowals are transparent-to-the-subject's-state, regardless of whether what is avowed is a phenomenal or intentional state, whether the avowal explicitly specifies some worldly matter or object outside the subject, or whether the avowal concerns a state that is itself rationally evaluable. The transparency-to-the-world of self-attributions of, specifically, *beliefs* falls out as but a special case or symptom of a broader phenomenon: the expressive character of all avowals. If asked (or when considering) whether you believe *p*, you will normally directly attend to whether *p* is *to be believed*. We can think of this as a way of putting yourself in a position to give direct voice to your (first-order) belief, which is what the neo-expressivist account says you do when avowing. You assess whether things are as the proposition says, and then simply pronounce on the truth of the proposition, though you are using a self-ascriptive expressive vehicle. But even unprompted, spontaneous pronouncements—such as “I’d like some tea,” or “I’m wondering what time it is”—that are not preceded by ‘direct consideration of the world’ can still be seen to partake in the expressive transparency of avowals.<sup>25</sup> It is this expressive transparency that separates avowals from both third-person and ‘alienated’ first-person mental-state attributions. For on the neo-expressivist account, a person’s avowal is not the upshot of an observation of, or even reflective attendance to, a state of herself; nor does it represent the result of an inference from a state of the world that she recognizes. Instead, an avowal is an upshot of the subject *speaking from* a present state of mind. As such, avowing can be seen as constituting a kind of ‘engaged awareness’ of our first-order mental states, as Boyle emphasizes, though on the neo-expressivist account this is not *reflective* awareness.<sup>26</sup> Unlike the subject who issues a report about the mental state of another individual, or the subject who issues an ‘alienated’ mental-state self-report, the avowing subject in effect embraces the first-order state that she self-attributes through her avowal, at least in the sense that she speaks from this state, giving it voice.

On the neo-expressivist proposal, it’s the expressive character of avowals, rather than any epistemic basis on which they are made (or the reasons for which the relevant self-beliefs are acquired or held) that explains avowals’ remarkable epistemic security. As expressive acts, avowals—like non-self-ascriptive expressions (including natural expressions), and unlike evidential reports (whether third- or first-person)—are indeed groundless, and *for that reason* protected from correction and demands for reasons or justification. However, insofar as avowals use as expressive vehicles truth-evaluable sentences that s-express self-ascriptive propositions, they are importantly different from other kinds of expressions of the relevant states. For, as regards avowals, we can sensibly raise questions concerning their epistemic status—we can ask, for example, what (if anything) qualifies them as items of (privileged) knowledge—a question to which we’ll return below.<sup>27</sup> But, contrary to the presupposition highlighted earlier, the neo-expressivist account does not suppose that providing an answer to this question about

avowals' epistemic status—by identifying positive, epistemically good-making features that render the relevant self-beliefs especially knowledgeable—is either necessary or sufficient for explaining avowals' distinctive security. On the contrary, the neo-expressivist account maintains that the latter explanation is to be given by appeal to avowals' expressive character; the explanation of what renders avowals instances of (privileged) self-knowledge is to be provided separately.

## 4. In Virtue of What Do Our Avowals Constitute Genuine Knowledge of Our Own States of Mind?

The neo-expressivist view of the epistemic security of avowals that we summarized in the previous section is compatible with a range of different accounts of what makes certain of our mentalistic self-ascriptions epistemically privileged in the sense of being especially likely to constitute knowledge.<sup>28</sup> However, we are here interested in exploring the possibility of a marriage between the neo-expressivist view of the security of avowals and a particular epistemic account of what makes our avowals knowledgeable when they are. To this end, we'll outline a *virtue-theoretic account* of self-knowledge. This account of self-knowledge is attractive in its own right, at least insofar as it inherits the merits virtue epistemology has in general.<sup>29</sup> But we'll aim to show, more specifically, how pairing the neo-expressivist account of avowals' security with a virtue-theoretic account of self-knowledge helps to both illuminate and make even more plausible the virtue theorist's explanation of what renders our avowals knowledgeable when they are.

As noted earlier, the neo-expressivist account of avowals' security does *not* appeal to a special epistemic basis on which avowals rest. Indeed, unlike both types of accounts we considered earlier, the neo-expressivist account is designed to explain—rather than explain away—avowals' groundlessness, by exploiting epistemic similarities between avowals and other expressive acts. Unlike traditional expressivist views, the neo-expressivist does not vindicate groundlessness by *assimilating* avowals to non-truth-evaluable expressions; rather, it locates the similarities between avowals and non-linguistic expressions in *acts* of avowing, while preserving the truth-evaluability of avowals as products. This means that neo-expressivism is able to remove a major *conceptual* barrier to regarding avowals as potential candidates for things we genuinely know. For avowals, according to neo-expressivism, are genuinely truth-evaluable self-ascriptions; they represent self-beliefs concerning which we can raise epistemological questions.<sup>30</sup> But this still leaves the neo-expressivist with a certain challenge: to explain

how, despite being epistemically groundless, avowals can nevertheless represent a kind of *epistemic achievement*.<sup>31</sup> Put differently, if it is to avoid deflationism, neo-expressivism must be paired with an explanation of what renders the second-order beliefs that our avowals express especially apt to constitute knowledge.

To anticipate, we will be arguing that these second-order beliefs are not ones that are *formed* on this or that epistemic basis, or *acquired* in response to this or that reason. In this sense, they could be regarded as beliefs one simply ‘finds herself with.’ Nonetheless, these second-order beliefs represent a genuine epistemic achievement. This is so because our having the relevant second-order beliefs is made possible by certain distinctive features of our psychological makeup in virtue of which we are in an especially privileged epistemic position when it comes to matters that concern our occurrent mental states. More specifically, our self-beliefs are especially apt to constitute knowledge courtesy of the fact that they manifest a kind of competence in belief possession, and this competence is grounded in the same features of our psychological makeup that enable us to express our states of mind by avowing.

We begin by drawing upon a general characterization of the notion of epistemic achievement adopted by virtue epistemologists. According to virtue epistemology, a subject’s belief that *p* constitutes a genuine epistemic achievement (and so, in the good case—when the belief is in fact true—constitutes an item of knowledge) just in case the subject manifests an *epistemic competence* in believing that *p*. And a cognitive capacity constitutes an *epistemic* competence if exercising that capacity reliably results in the subject’s having beliefs that achieve belief’s constitutive aim, goal, or standard of correctness (whatever that is taken to be).<sup>32</sup>

If we are to deploy a virtue-theoretic to explain self-knowledge, then our task is this: we must identify the capacity avowers exercise in possessing the self-beliefs their avowals express which, at least in the normal case, renders these self-beliefs especially likely to achieve belief’s constitutive aim.<sup>33</sup> Put differently, the central question for a virtue-theoretic account of self knowledge is:

What makes it the case that the second-order beliefs about our first-order states of mind that we ‘find ourselves with’ when avowing manifest epistemic competence, and so exhibit epistemic virtue?

We suggest that the neo-expressivist account of avowals’ security points to a certain way of answering this question. That account explains so-called first-person authority by appeal to our distinctive capacity to give voice to our mental states using self-ascriptive vehicles of expression (i.e., to avow). And our proposal is that we can see what renders avowable self-beliefs epistemically competent (that is, epistemicall virtuous and so

especially apt to constitute knowledge) by attending to whatever it is about our psychological makeup that (at least in the normal case) renders us capable of expressing our first-order states of mind by avowing. Thus, we suggest that the very same features of our psychological constitution simultaneously underwrite *both* the expressive capacity we manifest when we avow our first-order states of mind *and* the epistemic competence we manifest in holding second-order beliefs about these first-order states of mind.

On a virtue-theoretic picture, the epistemic status of any belief is a function of whether and how the belief manifests the believer's epistemic competence. It is the fact that a subject's having the belief manifests the believer's epistemic competence that *makes it the case* that the belief constitutes a genuine instance of knowledge. For the virtue theorist, then, a particular class of beliefs—perceptual beliefs, for example—will be especially apt to constitute knowledge if we are equipped with (and easily deploy) a set of domain-specific competences for possessing and regulating beliefs in this class. Thus, if our beliefs about domain D typically (i.e., in the normal case) manifest some domain-specific epistemic competence(s) that we have, then beliefs about D will be especially likely to constitute knowledge.<sup>34</sup>

Now, it can seem unquestionable that we exercise a kind of epistemic competence routinely and with ease in the domain of avowable self-belief. After all, it seems pre-theoretically clear that such beliefs constitute genuine knowledge. And if we take appearances at face value here, then, according to the virtue theorist, such beliefs must manifest *some* kind of epistemic competence. However, it is far from clear *in what* our epistemic competence to form and sustain avowable self-beliefs might consist.

We think that the neo-expressivist view points the way forward here. For it exposes a distinctive aspect of our psychological makeup in which our epistemic competence in the domain of avowable self-beliefs may reside. Specifically, we suggest that our capacity for *self-ascriptive expression* of our first-order mental states—the capacity that we exercise successfully in the normal case when we *avow* our first-order mental states—points to the psychological 'seat' of a distinctive *epistemic* competence the exercise of which our knowledgeable beliefs about these mental states typically manifest.

Neo-expressivism highlights the fact that certain features of our psychological constitution make it the case that we are capable of expressing our first-order mental states by avowing. And our proposal is that whatever features of our psychological makeup enable us to express our first-order mental states by avowing (thereby underwriting our expressive competence to avow) *also* ensure that the self-ascriptive beliefs we 'find ourselves with' regarding our present mental states manifest a domain-specific epistemic competence. The psychological 'seat' of the self-ascriptive expressive

capacity we exercise in avowing our first-order mental states is *also* the psychological ‘seat’ of our epistemic competence in the domain of mentalistic self-belief. If this is right, then it is the fact that a subject has the capacity to avow a particular occurrent mental state, M, that guarantees that her belief that she is in M represents a genuine epistemic achievement.

Even in the absence of a complete account of the features of our psychological constitution in virtue of which we are capable of expressing our mental states by avowing, there is good reason to think that a subject whose psychology exhibits those features must *also* be in a position to believe that she is in M in a manner that manifests her epistemic competence in the domain of self-belief. The reason, we suggest, is that the cases in which a subject’s expressive capacity to avow M suffers are precisely those cases in which the subject is not in a position to manifest the kind of epistemic competence that she might otherwise manifest in believing that she is in M. And this constitutes strong evidence that the psychological ‘seat’ of the relevant expressive capacity (i.e., the capacity to express first-order states of mind by avowing them) is also the psychological ‘seat’ of a distinctive epistemic competence.

In general, one can use the ways in which our ability to exercise a certain psychological capacity can be compromised, manipulated, or impaired to mark out, if only indirectly, the particular features of our psychological constitution which serve to ground that capacity. Our capacity to express first-order states of mind by avowing them is no exception. Both in everyday life and in the context of psychological research, it is easy to find cases in which subjects self-attribute a state of mind they are not in or are barred from avowing some state of mind that they are in. Such failures are thoroughgoing expressive failures: in paradigm cases where subjects fail to produce true avowals, they *also* fail to produce appropriate *non-self-ascriptive expressions* of their states of mind. So, for example, a subject who falsely avows “I don’t believe women are less qualified for high-ranking positions” (thereby revealing a certain self-ignorance) would *also* be disposed to deny the first-order, non-self-ascriptive claim “Women are less qualified for high-ranking positions....” Similarly, a subject not in pain who avows “I’m in pain” as a result of being primed by the sight of an approaching drill at the dentist’s would *also* be disposed to wince.<sup>35</sup> Moreover—and this is the crucial point for our purposes here—both sorts of expressive failure also, and inevitably, involve a failure of self-knowledge. When a subject avows a state of mind that she is not in, the subject’s avowal expresses a false second-order belief (and so a belief that fails to constitute knowledge) about her own state of mind. And when a subject exhibits an inability to avow some state of mind that she is in, we don’t credit the subject with a second-order belief attributing the relevant first-order state of mind (and so with non-alienated self-knowledge).

What this means is that there is a very tight connection between failures of self-knowledge, on the one hand, and failures to be appropriately engaged with one's states of mind so as to be able to give them voice (whether by avowing them *or* by producing various non-self-ascriptive expressive behaviors), on the other hand. When a subject's psychology is compromised, manipulated, or impaired such that her expressive capacity suffers, her epistemic competence to achieve self-knowledge suffers as well. And vice versa. Cases in which a subject fails to have ordinary, basic self-knowledge are cases in which the subject is not in a position to *avow* her states of mind (though she may still be able to self-attribute those states through 'alienated' self-reports). This, we submit, provides very good evidence that our capacity to express ourselves by avowing and our epistemic competence for possessing knowledgeable mental-state self-beliefs are grounded in the very same features of our psychological makeup; they share a psychological 'seat'.

If all this is right, then the fact that S has the capacity to express M by avowing ensures that S is capable of exercising a distinctive epistemic competence in believing that she is in M, and so S is especially likely to know that she is in M. What is involved in exercising this distinctive epistemic competence? Crucially, if our account of self-knowledge is to accommodate the groundlessness of avowals, then what it takes to exercise this epistemic competence cannot be understood on the model of what is involved in exercising an inferential or perceptual competence. This is because both inferential and perceptual competences make an essential appeal to the way in which the virtuous epistemic subject *moves from substantive, positive reasons for belief to belief itself*. But cashing out our epistemic competence in the domain of mental-state self-belief by appeal to the way in which the virtuous epistemic subject moves from positive reasons for belief to belief itself amounts to denying groundlessness. Such a model of our epistemic competence in the domain of mental-state self-belief would credit someone with *knowledge* that she believes that p courtesy of the fact that she has *formed* her second-order belief *on the basis of* positive reasons for that belief. So a virtue-theoretic account of self-knowledge that aims to accommodate groundlessness must explicate the relevant epistemic competence(s) in a way that does *not* appeal to subjects' movement from positive reasons for self-beliefs to the self-beliefs themselves.

Can the virtue-theoretic account that we have begun to sketch make good on this demand? We are optimistic that it can. First, notice that, on the view that we are developing here, it is *not* that the expressive capacity to avow is *itself* a kind of epistemic competence via which we arrive at self-ascriptive beliefs. It is not as though we *base* our self-beliefs on the outputs of the exercise of our expressive capacity. (To suggest that would be (1) to fall back into a rather crude form of behaviorism; to wit: we witness our avowals and form our self-beliefs on that basis,<sup>36</sup> and (2) to treat neo-expressivism, by

itself, as a view about what makes our avowals knowledgeable, which it is not.) But more generally, the virtue-theoretic account that we have sketched in this section need not model our epistemic competence in the domain of self-belief as a kind of rational transition from reasons for belief to belief itself (i.e., from positive grounds for belief to belief based on those grounds). To be sure, some epistemic competences (e.g., perceptual or inferential competences) are fruitfully, and perhaps unavoidably, understood in this way. But some, we think, are not. Consider, for example, our knowledge of very basic mathematical facts (e.g., that  $2 + 2 = 4$ ) or of obvious, apparently analytic truths (e.g., that all bachelors are unmarried). Or consider the chicken-sexer's belief that the chick in front of her is, as it might be, male. It seems that the believer in these cases (as we put it earlier) simply 'finds herself with' the relevant beliefs. And, we think tellingly, one would be hard-pressed to identify a person-level state that stands in the right reason-giving relation to the believer's belief that  $2 + 2 = 4$ , that all bachelors are unmarried, or, in the case of the chicken-sexer, that a particular chick is male. But, importantly, these beliefs seem to have a very different epistemic status from, say, wild guesses, or beliefs inserted into the believer's mind through hypnosis or via other arbitrary means. And this difference is easily explained by the fact that, in the former type of cases, there is undeniably some kind of epistemic competence (although clearly not an inferential competence) the exercise of which is manifest in the possession of the belief the subject finds herself with. Not so in the latter type of cases. The epistemic competence that the chicken-sexer exercises in believing that the chick in front of her is male/female, for example, is not appropriately understood as involving a disposition to move from some positive reason for a certain sort of belief to belief based on that reason. Nonetheless, it plausibly constitutes a kind of epistemic competence that can underwrite genuine knowledge. And so, we suggest, there is good reason to accept that the distinctive epistemic competence that we paradigmatically manifest in mentalistic self-belief can be coherently construed, on this model, in a way that vindicates the groundlessness of avowals.

## 5. Some Concluding Remarks

We began our discussion with a puzzle concerning self-knowledge: our avowals appear to be at once epistemically groundless and epistemically privileged. We have argued that recent attempts to capture what is epistemically distinctive about the self-beliefs that avowals represent by appealing to the feature of transparency require treating avowals' groundlessness as merely apparent. (The inferentialist treats avowals as expressing beliefs that possess their positive epistemic status in virtue of being inferentially grounded in reasons for the self-attributed beliefs. Whereas the reflectivist treats them as

expressing beliefs that are rationally justified in virtue of the fact that they make explicit what a subject already—if only tacitly—knows simply by *being* in the self-attributed state.) In contrast, our aim was to provide an account that both vindicates and explains avowals' groundlessness without denying that avowals represent second-order beliefs that are especially apt to constitute knowledge. Our proposal here marries the neo-expressivist explanation of avowals' groundlessness with a virtue-theoretic account of self-knowledge. We've argued that the neo-expressivist's non-epistemic account of avowals' security helps to locate a distinctive epistemic competence whose exercise can render our avowable self-beliefs—beliefs about our own mental states that we simply 'find ourselves with'—epistemically virtuous: there is at least indirect evidence that the relevant epistemic competence shares a psychological seat with the expressive capacity we exercise when avowing (which capacity, according to neo-expressivism, explains avowals' distinctive security). A comprehensive characterization of the psychological seat of this epistemic competence, and so a complete account of what it takes for a subject to exercise the competence, awaits further investigation. However, we hope to have shown here that a virtue-theoretic line of investigation will be well worth the effort of any theorist of self-knowledge who wishes to preserve both the groundlessness of avowals and the idea that they represent beliefs that are especially likely to constitute genuine knowledge of our first-order mental states.<sup>37</sup>

## Bibliography

- Bar-On, Dorit (2015). "Expression: Acts, Products, and Meaning," forthcoming in *Meaning without Representation: Essays on Truth, Expression, Normativity, and Naturalism* (Steven Gross, et. al. eds.), Oxford, UK: Oxford University Press.
- Bar-On, Dorit (2012). "Externalism and Skepticism: Recognition, Expression, and Self-Knowledge," in *Self-Knowledge and the Self* (A. Coliva, ed.), Oxford, UK: Oxford University Press.
- Bar-On, Dorit (2010). "Avowals: Expression, Security, and Knowledge: Reply to Matthew Boyle, David Rosenthal, and Maura Tumulty," *Acta Analytica* 25: 47–64.
- Bar-On, Dorit (2009a). "First-Person Authority: Dualism, Constitutivism, and Neo-Expressivism," *Erkenntnis* 71: 53–71.
- Bar-On, Dorit (2009b). Review of Simon Evnine, *Epistemic Dimensions of Personhood*, *Philosophical Books* 50 (1): 1–14.
- Bar-On, Dorit (2004). *Speaking My Mind: Expression and Self-Knowledge*, Oxford, UK: Clarendon Press.

- Bar-On, Dorit and Long, Douglas (2003). "Knowing Selves: Expression, Truth, and Knowledge," in *Privileged Access: Philosophical Accounts of Self-Knowledge* (Brie Gertler, ed.), Ashgate Epistemology and Mind Series, 2003, 179–212.
- Boghossian, Paul (1989). "Content and Self-Knowledge," *Philosophical Topics* 17: 5–26.
- Boyle, Matthew (2011). "Transparent Self-Knowledge," *Aristotelian Society Supplementary Volume* 85 (1): 223–241.
- Boyle, Matthew (2010). "Bar-on on self-knowledge and expression," *Acta Analytica* 25 (1): 9–20.
- Byrne, Alex (2011). "Transparency, Belief, Intention." *Aristotelian Society Supplementary Volume* 85 (1): 201–221.
- Byrne, Alex (2005). "Introspection," *Philosophical Topics* 33 (1): 79–104.
- Carruthers, Peter (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press.
- Coliva, Annalisa (2012) in *The Self and Self-Knowledge* (Coliva, ed.). Oxford University Press: 212–242.
- Evans, Gareth (1982). *The Varieties of Reference*. Oxford University Press.
- Fernández, Jordi (2003). "Privileged Access Naturalized," *Philosophical Quarterly* 53: 352–372.
- Fricker, Elizabeth (1998). "Self-Knowledge: Special Access vs. Artefact of Grammar—A Dichotomy Rejected," in Cynthia MacDonald, et. al. (eds.) *Knowing Our Own Minds*. Oxford University Press: 155–206.
- Gallois, André (1996). *The Mind Within, the World Without*. Cambridge University Press.
- Gertler, Brie (2011). *Self-Knowledge*. Routledge.
- Greco, John (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.
- Moran, Richard A. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.
- Nolfi, Kate (2015) "How to be a Normativist about the Nature of Belief." *Pacific Philosophical Quarterly* 96 (2): 181–204.

Sellars, Wilfrid (1969). "Language as Thought and as Communication," *Philosophy and Phenomenological Research* 29: 506–527.

Sosa, Ernest (2009). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford University Press.

Valaris, Markos (2011). Transparency as Inference: Reply to Alex Byrne. *Proceedings of the Aristotelian Society* 111 (2pt2): 319–324.

Zagzebski, Linda (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge University Press.

### Notes:

(<sup>1</sup>) For discussion of avowals' epistemic security and an account of it see, *inter alia*, Bar-On (2004).

(<sup>2</sup>) See, for example, Gallois (1996), Moran (2001), Fernandez (2003), Byrne (2005).

(<sup>3</sup>) Inner-sense theories, recall, pin the security of avowals and 'first-person authority' to the high degree of reliability—even if not Cartesian infallibility—of our faculty of inner perception; and acquaintance theories pin it to the existence of an unmediated relation of direct acquaintance we have to our own current states of mind. For a survey, see Gertler (2011: Ch. 5).

(<sup>4</sup>) For discussion, see Bar-On (2004: Chs. 4, 8). Following Evans, it seems preferable to speak of transparency-to-the-world as a feature of (some) *self-ascriptions of beliefs* and other states of mind (that is, second-order judgments on our first-order mental states), rather than speaking—as do some recent authors—of our first-order *beliefs* (and other states) *themselves* as enjoying transparency. See Bar-On (2009a).

(<sup>5</sup>) This reading of Evans is presented in Bar-On (2004, Ch. 4); see also, Bar-On (2009b).

(<sup>6</sup>) In (2005: 94), Byrne talks of the 'epistemic rule':

BEL If p, believe that you believe that p.

Which he formulates to fit epistemic rules of the general form

R If conditions C obtain, believe that p.

(He considers, for example,

DOORBELL If the doorbell rings, believe that there is someone at the door.

Or

NEWS If the Weekly News reports that p, believe that p.)

(<sup>7</sup>) The schema is intended to capture a ‘cognitive transition’ that subjects make, or can make, which is not intended to be an explicit, conscious inference. But see Boyle (2011: 230–231).

(<sup>8</sup>) Also crucial in Byrne’s formulation here is that recognition entails knowledge. Thus, the fact that a subject recognizes that the antecedent obtains entails that the subject knows (and so also that the subject believes) that it obtains. However, even a failed attempt to follow BEL will often produce knowledge—for even if p is false (and hence not recognized or known), such attempts usually involve the belief that p.

(<sup>9</sup>) Note, however, that this may be a peculiarity of the case of belief. See section 3.

(<sup>10</sup>) Byrne understands *recognition* in terms of ‘cognitive contact’ with the relevant state of affairs. This is what helps explain why the rule BEL is *self-verifying*. However, there are cases in which one recognizes p, but also affirms that one doesn’t believe p (or believes not-p) based on, for example, a therapist’s analysis, or interpretive self-analysis. So cognitive contact with p isn’t sufficient for transparency; one must also *attribute* the belief ‘from a 1st-person perspective’. It is not clear how Byrne can accommodate this fact. (And see note 19 below.)

(<sup>11</sup>) See Byrne (2005: 96–98) and (2011: 206f).

(<sup>12</sup>) Byrne extends his account to cover self-attributions of mental states other than belief. Just to give the flavor of the extended account, Byrne (2013) also proposes:

INT I will  $\emptyset$

-----

I intend to  $\emptyset$

(<sup>13</sup>) See Byrne (2011) for relevant references and objections. ‘Detectivist’ views pin self-knowledge to the subject’s ability to *detect* reliably (perhaps even infallibly) her own mental states by, from example, being able to introspect their presence.

(<sup>14</sup>) Consider: I sometimes *suppose* *p* just for the sake of argument (say, in a conditional proof), or toward a *reductio*. In such cases, it would be clearly wrong for me to move from (my supposition that) *p* to (an affirmation of) “I believe that *p*.” But (on pain of circularity) it seems that Byrne cannot restrict the inference to cases in which I know that my starting point is my genuinely believing that *p*. For relevant discussion, see Valaris (2011).

(<sup>15</sup>) Compare Bar-On (2004: 95ff.).

In the case of a rule like INT (see fn. 12), the misguided idea is that “I might infer propositions about my present intentions from blank future propositions about myself from a blank future proposition about myself, as if I must conclude my own commitment to  $\emptyset$  from an unaccountable inkling about what I will in fact do” (2011: 234).

Boyle thinks that Byrne’s inferentialist approach to doxastic transparency faces a certain dilemma: “It must either represent the subject as drawing a mad inference, or else must admit that her real basis for judging herself to believe *P* is not the sheer fact that *P*, but her tacit knowledge that she believes *P*. The second horn of this dilemma should be unacceptable to Byrne: embracing it would mean giving up on his project.” (Boyle 2011: 234).

(<sup>16</sup>) These are features that Boyle’s account shares with ‘constitutivist’ views (which take self-knowledge to be a constitutive feature of mental states). For discussion of constitutivism and references, see Gertler (2011: Ch. 6), Coliva (2012), and Bar-On (2009a).

(<sup>17</sup>) For some further discussion, see Bar-On (2009a), (2010), and (2015).

(<sup>18</sup>) Notably, transparent consideration of outward phenomena can issue in self-ascriptions that are ripe for being caught in Moorean absurdities. But here too we should be wary of excessive focus on belief. The following all seem anomalous in much the same way as the Moore belief sentences: “I’m finding this meeting really exciting, but it’s very boring”/ “I am not pleased to see you, but it’s so great to see you”; “Tea please! But I don’t want any tea”; and also “Brrr! It feels hot in here” and “[Agonized expression]; I feel so happy.” It does not seem right to capture what is commonly anomalous across all these cases in terms of what is required by rationality (unless we understand ‘rationality’ very loosely). (Moreover, *even focusing on belief*, one must keep in mind that Moorean conjunctions *can* be rationally produced or entertained; I *can* sincerely affirm *p*, yet have independent reasons (based on evidence, inference, testimony, etc.) for thinking that I don’t believe it, or that I actually believe the opposite. So the connection to rationality is far from straightforward even in the case of belief.) (For discussion, see Bar-On 2009b.)

(<sup>19</sup>) See also, *inter alia*, Bar-On (2010), and (2012).

(<sup>20</sup>) A-expression is a three-place relation: an agent J a-expresses mental state M by using expressive means or vehicle E, where E can be bodily demeanor, facial expression, or gesture, whether natural, culturally acquired, or even idiosyncratic; it can *also* be a bit of verbal behavior.

(<sup>21</sup>) We can here set aside, for the most part, what Sellars (1969) calls (misleadingly) “expression in the *causal* sense”—for example, nonvoluntary, uncontrolled facial expressions or gestures that reveal one’s state of mind. This is because the expressive behaviors relevant to our concerns here—avowals—are *not* nonvoluntary or reflexive bodily happenings, but rather things that are *done by an individual* (as opposed to a sub-system, or module, within the individual), over which the individual exercises a certain kind of central, executive control. See Bar-On (2004: 216f., 249ff., 289, 315).

(<sup>22</sup>) In general, acts of expressing one’s state of mind can be said to deploy expressive vehicles that are in some sense *designed to show* that state. Natural expressions such as growling, teeth baring, gaze shifting, smiles, and so on, are designed by nature to show the relevant states to suitably endowed receivers. But the idea is arguably applicable to acquired or learned expressions, including linguistic ones. For discussion, see, for example, Bar-On (2004: Ch. 7), (2010).

(<sup>23</sup>) Contrast: “Rain would be great!” or: “Oh for some rain!,” which may equally serve to a-express a subject’s wish for rain, but without *naming* the state. (Note that, in the case of verbal expressions, the information regarding the expressed state need not be revealed through features of nonverbal expressive behavior; it is made available through the linguistic vehicle used.)

(<sup>24</sup>) See Bar-On (2004: 264ff., 310ff).

(<sup>25</sup>) The commissive aspect of transparent self-attributions of belief (and other reflective states) emphasized by reflectivist views can also be seen as a consequence of avowals’ expressive character. Insofar as avowing a belief (as opposed to merely reporting it) serves to give direct expression to the self-ascribed belief itself, one incurs commitment to the truth of one’s first-order belief. (See Bar-On 2004: 135ff., 318f., and Bar-On 2009b.)

(<sup>26</sup>) Boyle (2011). See also Moran (2001).

(<sup>27</sup>) Thus, although the neo-expressivist account borrows from traditional avowal expressivism the insight that avowals’ distinctive security is due to their similarities to other expressive acts, it departs from the traditional account in highlighting the fact that, like various mental and non-mental descriptive reports, avowals use *expressive vehicles*—

sentence- or thought-tokens—that are semantically complex and are truth-evaluable. For a full development of this account and an explanation of the strong *presumption of truth* governing avowals, see Bar-On (2004: chs. 6–8).

(<sup>28</sup>) See Bar-On and Long (2003) and Bar-On (2004: Ch. 9).

(<sup>29</sup>) See Zagzebski (1996), Sosa (2009), Grecco (2010).

(<sup>30</sup>) This means that avowals express *both* the self-attributed states and the avowers' higher-order belief that she is in the state. For discussion of the merits of this 'dual expression' thesis, see Bar-On (2004: 307–310 and 366).

(<sup>31</sup>) For discussion of 'epistemic achievement,' see Boghossian (1989) and Fricker (1998).

(<sup>32</sup>) Most virtue epistemologists construe this aim in terms of truth (for example, see the authors mentioned in note 29). However, one of us has argued elsewhere that the aim of belief must be understood in terms of belief's distinctive functional role in believers' mental economies, rather than in terms of truth (Nolfi 2015). Our proposal here is neutral between these different ways of characterizing the aim of belief.

(<sup>33</sup>) As noted earlier (note 30), we take it that when you avow, you express *both* the first- and the higher-order state. That means you already have the self-belief—avowing is NOT the act/process/method that gives rise to or issues in this belief.

(<sup>34</sup>) Notice that it is somewhat misleading to think of avowals as expressing beliefs that belong to a specific *domain*. The relevant beliefs cannot simply be individuated in terms of their content, or subject-matter. We take it to be an advantage of the account that we go on to outline that it seems especially well placed to capture what is distinctive of the relevant beliefs: they are avowable. And that the relevant mental states have this distinctive feature allows us to make sense of the idea that we might be able to deploy specific, targeted competences in coming to know those of our first-order mental states that have the feature in question, competences that we cannot deploy elsewhere.

(<sup>35</sup>) For discussion and references, see Bar-On (2004: 320ff.), (2009a).

(<sup>36</sup>) New versions of this Rylean view have been recently proposed. See, for example, Carruthers (2011).

(<sup>37</sup>) We wish to thank Ram Neta, Carol Voeller, and an anonymous referee for Oxford University Press for reading and commenting on earlier versions of this piece.

**Dorit Bar-On**

Dorit Bar-On is a Professor in the Department of Philosophy at the University of Connecticut

**Kate Nolfi**

Kate Nolfi, Department of Philosophy, University of Vermont

