

Expressing One's Mind

David M. Rosenthal

Received: 30 November 2009 / Accepted: 1 December 2009 / Published online: 7 January 2010
© Springer Science+Business Media B.V. 2009

Abstract Remarks such as ‘I am in pain’ and ‘I think that it’s raining’ are puzzling, since they seem to literally describe oneself as being in pain or having a particular thought, but their conditions of use tend to coincide with unequivocal expressions of pain or of that thought. This led Wittgenstein, among others, to treat such remarks as expressing, rather than as reporting, one’s mental states. Though such expressivism is widely recognized as untenable, Bar-On has recently advanced a neo-expressivist view, on which such remarks exhibit characteristics of both expressions of mental states and reports of those states. I argue against any attempt to see such remarks as both reporting and expressing the same mental states, and that a correct account rests on distinguishing the truth conditions of such remarks from their conditions of use.

Keywords Expressivism · Wittgenstein · Sellars · Moore’s paradox · False belief

1 Introduction

Remarks such as ‘I am in pain’ and ‘I think that it’s raining’ present opportunity for reflection and theory. Ostensibly such remarks report what one feels or thinks, but we do not in conversation treat these remarks as we do ordinary reports. If I ask you about the weather and you say, ‘I think it’s raining,’ I can’t complain that you told me just about your thoughts, and not about the weather. It is often held, moreover, when we do take such remarks as revealing the speaker’s mental states, those remarks are not subject to the kind of challenges that are in place with ordinary reports. Indeed, such remarks are often taken to exhibit some kind of epistemic privilege, and some have even maintained that one cannot be wrong when one says such things.

For these and related reasons, Wittgenstein held that in saying such things, we do not actually report the mental states we seem to mention; rather, we simply express

D. M. Rosenthal (✉)
Program in Philosophy, CUNY Graduate Center, 365 Fifth Avenue, New York,
NY 10016-4309, USA
e-mail: davidrosenthal@nyu.edu

those states. Saying ‘I am in pain’ expresses, rather than reports my pain (Wittgenstein 1953, Part I, §§244, 310). It is in that way on par with crying out or saying ‘ouch’. Similarly, Wittgenstein maintained, saying ‘I think it’s raining’ expresses my thought that it’s raining, rather than reporting that thought. Thus he writes: ‘[T]he statement “I believe it’s going to rain” has a meaning like, that is to say a use like, “It’s going to rain”’ (Wittgenstein 1953, IIx, p. 190). On this expressivist view, sentences that ostensibly report one’s own beliefs, like those that ostensibly report one’s own sensations, would then really just express those beliefs and sensations. It is this general idea that Douglas Gasking (1962) sought to capture by dubbing such remarks *avowals*.

An apparent advantage of this view is that it seems to explain the sense we have that these apparent self-ascriptions of mental states have some epistemic privilege. Saying ‘ouch’ plainly expresses one’s pain; it does not report it. So if I cry out or say ‘ouch’, it makes no sense to tell me that I’m wrong. The same will hold for saying ‘I am in pain’ if in saying that I simply express the state in question, and don’t actually report it. We can in this way explain the appearance of epistemic privilege without positing any special access to our inner states.

As it stands, however, this picture won’t do. As P. F. Strawson (1959) and Hilary Putnam (1968) among others have stressed, whatever our conversational habits as regards such remarks, first-person remarks about one’s own mental states interact with other speech acts in all the ordinary, logical ways. When you say of me ‘You aren’t in pain’ and ‘You don’t think it’s raining’, your remarks contradict what I assert if I say ‘I am in pain’ and ‘I think it’s raining’. By contrast, when I say instead ‘It’s raining’, I express my thought that it’s raining, and in saying ‘ouch’ I express my pain. So those speech performances, if ‘ouch’ does indeed count as a kind of speech, do not literally contradict your saying that I’m not in pain or that I don’t think it’s raining.

All this makes it tempting to see whether we can in some way do justice both to the reporting role such remarks evidently have as well as to the ways these remarks appear to diverge from ordinary reports. Perhaps we can somehow see such remarks as combining features of ordinary reports with features of uncontroversial expressions of mental states, such as one’s saying ‘ouch’ or saying simply ‘It’s raining’.

2 Bar-On’s Neo-Expressivism

The Wittgensteinian view that speech acts such as ‘I am in pain’ and ‘I think that p ’ express the pain and the thought that p , respectively, and do not literally self-ascribe those states, seems to account for the sense we have that such speech acts have a special epistemic status. But that view runs up against the difficulty that those speech acts then fail to interact with other, third-person speech acts in characteristic conversational and logical ways. So it’s tempting to see whether there is a way to see such speech acts as sharing features both of expressions of mental states and literal self-ascriptions of those states.

Dorit Bar-On (2004) has recently developed a more detailed, sustained argument that these speech acts do have features both of ordinary reports and of purely

expressive remarks. In what follows I critically examine Bar-On's challenging defense of that view.

Bar-On refers to the difficulty noted in §I for Wittgenstein's purely expressive view as *semantic continuity*. Semantic continuity holds that spontaneous speech acts, such as 'I am in pain' and 'I think that p ', have the semantic properties that they seem on the surface to have. In particular, they are semantically related in the obvious ways to such speech acts as 'You are in pain', 'I was in pain', 'George thinks that p ', and 'I used to think that p '. Bar-On refers to spontaneous, first-person speech acts such as 'I am in pain' and 'I think that p ' as *avowals*, and for convenience I'll follow that usage in the remaining discussion.

Bar-On endorses semantic continuity, and sees it as decisive against the unmodified Wittgensteinian view, which she calls simple expressivism. But she also sees simple expressivism as having an important strength. Avowals, she insists, have a distinctive epistemic security in being 'protected from ordinary epistemic assessments' (20; cf. 123, 222), though avowals are not, she acknowledges, absolutely infallible or incorrigible. This epistemic security has no exact parallel with other types of speech act, and cannot be detached from ordinary avowals. Seeing avowals such as 'I am in pain' and 'I think that p ' as expressing the pain and the thought that p themselves explains, Bar-On urges, how and why these avowals are protected from such epistemic assessments. It's always appropriate to ask for reasons and raise challenges when somebody describes a mental state. Such responses are largely inappropriate, however, if one is instead 'giving voice to the [mental] condition' (263) in question.

On the standard alternative to Wittgensteinian expressivism, speech acts such as 'I am in pain' and 'I think that p ' describe, rather than express, the mental states they are ostensibly about. So advocates of such views typically seek to capture a measure of epistemic security for avowals in purely epistemological terms, by invoking an epistemic advantage that occurs when we say what mental states we are in, but is absent with other types of speech act. But Bar-On argues that there is no epistemic advantage that does justice to the epistemic security that avowals have, or even that distinguishes avowals from all other speech acts. These alternative views, she concludes, fail to capture the epistemic asymmetry between avowals and other speech acts.

How, then, might we capture that epistemic asymmetry without sacrificing semantic continuity? Bar-On's solution is a neo-expressivist view that draws both from the simple expressivism of Wittgenstein and from the descriptivist alternative. Like simple expressivism, Bar-On's neo-expressivism holds that when I say that I am in pain or that I think that p , my speech act expresses my pain or my thought that p . But neo-expressivism, like the descriptivist alternative, also holds that these speech acts serve literally to ascribe those states to myself. Avowals on this view 'are not protected by grammar' from being false (339). They are protected from epistemic challenge and 'strongly presume[d]' to be true simply because they are a case of speaking one's mind (311).

I am less confident than Bar-On that avowals have the kind of epistemic security she attributes to them. Bar-On gives little argument for the type of security she sees avowals as having, largely taking it to be an intuitive datum. And a descriptivist might well see the epistemic security she attributes to avowals as being somewhat

too conveniently keyed to what the expressivist and neo-expressivist can deliver. So I am not as convinced as Bar-On that the descriptivist cannot capture the security avowals actually do have by appeal to a mixture of epistemic and pragmatic considerations.

I'll say just a bit more about this in §VI. But my main concern about Bar-On's neo-expressivism has to do with its endorsement of the Wittgensteinian idea that avowals actually express, in a specific way I'll now turn to, 'the avowed [mental] condition' (307).

3 Expressing a State

Following Wilfrid Sellars (1969; cf. 1964), Bar-On distinguishes the expressing of mental states from the expressing of propositions (216). It's central to her view that each avowal expresses, in the mental-state way, the mentioned state, but also expresses, in a different, semantic way, the proposition that one is in that state.

On the descriptivist view, the speech acts that Bar-On calls avowals do not express in the mental-state way the state it mentions; rather, they each express a thought that one is in the relevant state—a thought that I have in another context (e.g., 2005) called a higher-order thought (HOT). Bar-On concedes that each avowal, in addition to expressing the proposition that one is in the state it mentions, may also express a mental judgment to that effect (307–310). But whatever the case about that, she insists that avowals express, in the mental-state way, the states that they explicitly mention.

Because expressing a mental state can be an intentional action, Bar-On relies on intentions to determine which state is expressed by an avowal. For a piece of behavior 'to express [a mental state] M in the [relevant] sense,' she writes, 'is to have M as one's reason for the relevant behavior' (309). And she takes this appeal to the reasons speakers have in issuing avowals to show that avowals such as 'I am in pain' and 'I think that p ' express the states that they mention. But it's not obvious that this appeal to reasons does actually support that conclusion.

On the reasons test, my speech act 'I am in pain' expresses the pain itself only if the pain is my reason for performing that speech act. But the pain itself is not my reason. My reason is something along the lines of my desire to let you know that I am in pain, not the pain itself. Similarly, my reason for my saying 'I think that p ' is not my thought that p itself, but my desire to let you know, or even just to express, that I think that p .

Reasons for doing things generally consist in some combination of beliefs and desires. My reason for knocking on the door is my desire to get you to open the door, together with my belief that knocking will achieve that result. My pain is not the kind of thing that would figure in my reasons for saying something.

Perhaps my awareness of being in pain would so figure. Presumably I wouldn't say 'I'm in pain' unless I thought I was in pain. And Bar-On does allow that the avowal, 'I am in pain', may well express not only the pain itself, but also my 'judgment' (307) that I am in pain. So perhaps she would also accept that it is that judgment, or HOT, which figures in one's reason for saying 'I am in pain'. I'll say a bit more, in my closing §VII, about such HOTs.

We sometimes speak of something as the reason for doing something when it's the occasion that motivates one's doing that thing. An explosion, for example, may be the reason for my running off. My pain can in that way be my reason for saying 'I'm in pain'; it's the event that provokes my behavior, in this case my speech act. But a pain's being my reason in this way doesn't support the idea that my speech act expresses the pain. Compare the explosion; my running off doesn't express the explosion.

The appeal to reasons to determine which state a speech act expresses does not in any case square with the ordinary way we describe speech acts as expressing mental states. If I sincerely say 'It's raining', my speech act is ordinarily taken to express my thought or belief that it's raining (cf. 308). That is, we take it to express a mental state with the same intentional content as the speech act and a mental attitude that corresponds to the speech act's illocutionary force. So it's natural to see the avowal, 'I am in pain', which Bar-On concedes has the semantic meaning that one is in pain, as also expressing a judgment, or thought, that has that content. I myself believe that it's this commonsense notion of expressing that Sellars sees as basic, rather than a notion that hinges on intentions or reasons (1969: §§IX–X). But whatever the case about that, some different notion of expressing must in any case be operative if the avowal, 'I am in pain', is to express the pain itself.

Sellars does describe what he calls an action sense of expressing, on which an individual expresses an intentional state by saying something. And Bar-On urges that avowals express the states they mention in this action sense. But Sellars argues that this action sense is constructed from, and so relies on, a more basic kind of expressing, on which speech acts constitute a kind of thinking out loud, and thereby express intentional states that have the very same content as the speech acts that express those states. Sellars calls this the causal sense of expressing, and argues that, though expressing in this way is a type of doing something, it is not on that account a voluntary action (1969: §IX). I believe Sellars is correct in seeing the action type of expressing as built on the causal type (1969: §X). If so, action expressing will follow causal expressing, so that the state a speech-act action expresses will have the same intentional content as that speech act itself.

If I say 'It's raining', I semantically express the proposition that it's raining. But Sellars (1964) stresses that this amounts only to the speech act's belonging to a set of speech acts that have a particular functional role. Causal expressing, however, is more demanding; I cannot causally express a state I'm not in, since a state that doesn't occur cannot cause anything.

If I utter 'It's raining' but have no such thought, my utterance may still count as expressive behavior, as Bar-On urges. But that means only that it is a token of a type that typically does express a corresponding intentional state (Rosenthal 2005, ch. 3, §5). Insincere speech, as Frege (1977, 8) noted, is pretend speech. If I utter 'It's raining' without the corresponding thought, I merely pretend to assert that it's raining; I pretend, that is, to be a person just like me except that I actually have that thought. The causal notion of expressing can thereby remain intact.

Bar-On urges that the avowal, 'I think that p ', semantically expresses the proposition that one thinks that p and expresses in the action sense the thought that p . But I think it's again better to follow Sellars in seeing such semantic expressing as itself also a construction from causal expressing. Speech acts of asserting 'It's

raining' have a particular functional role. The proposition that it's raining is, according to Sellars, an abstraction from particular speech acts, in whatever language, that have that role together with the intentional states that play the analogous role in one's mental economy (Sellars 1968, ch. 3; Rosenthal 2005, ch. 3). So to say that 'I think that p ' expresses a particular proposition is just to describe its semantic functional role in abstract terms.

Bar-On sees avowals (296), along with naturally expressive behavior such as wincing (272 ff.), as actually showing the expressed state to others, and as allowing others to see it. This goes well beyond Wittgenstein's commonsense observation that somebody's saying "'I believe' throws light on [that person's] state' (1953: II, x, p. 191). According to Bar-On, an 'observer perceives the state' another individual expresses, or in any case 'perceives [that individual's] being in the relevant state' (278). Bar-On even maintains that when expressive behavior has become second-nature, it actually 'becom[es] a characteristic ... component of the relevant state' (299). But it's implausible that expressive behavior, however characteristic or second-nature, is literally a component of the state expressed, or that we literally see the state, as against simply seeing, that is, concluding, that somebody is in that state. After all, the pains and thoughts can each occur without the expressive behavior, and conversely. The tie between state and expressive behavior is just not that close.

4 Moore's Paradox

Bar-On holds that avowals express the very states that they explicitly mention. But avowals also have the semantic content that one is in those states. And since avowals are not infallible, they may on occasion be false; even on Bar-On's view, the security involves only a strong presumption of truth (311). But when avowals such as 'I am in pain' and 'I think that p ' do fail to be true, there is no relevant pain or thought that p . So how in those cases can avowals express those very states?

Bar-On urges that even when a sincere remark that one's tooth hurts fails to be true, we can still see that remark as an avowal—as opposed to any other kind of speech act. Sincerely saying that one's tooth hurts is speaking one's mind; it is 'speaking directly from a present condition.' Even if there is no state for the avowal to express, it is 'an expressive act' (328).

But if there is no state for the avowal to express, the speech act's being an expressive act can mean only that it is the kind of speech act that ordinarily expresses such a state. And a uniform account is available, which has no need to make special provision for untrue avowals. All sincere speech acts involve speaking, in an intuitively direct way, from one's present mental condition. If I sincerely say 'It's raining', I speak, seemingly directly, from my thought that it's raining, and that's so even when my statement that it's raining isn't true. So the descriptivist will agree that even when my sincere remark that I'm in pain or that I think that p is not true, I still speak directly from the state I express. It's just that the state my avowal expresses is not the state I report; rather, it is a thought about that state, for example, a thought that I am in pain or a thought that I think that p . The descriptivist need not deal in any special way with cases in which one's sincere self-ascriptions fail to be true.

A related issue arises in connection with Bar-On's treatment of Moore's paradox. As Moore (1942: 543; 1944: 204) noted, one cannot coherently say things like 'It's not raining, but I think it is'. Bar-On acknowledges that the difficulty is not strictly semantic, since such sentences plainly are not contradictory; it may well not be raining even if I think it is. Still, one cannot coherently say such things. The problem is to explain why.

Bar-On explains it by arguing that, even though the two conjuncts do not express incompatible propositions, they do express 'two conflicting beliefs' (218). Plainly, the conjunct 'It's not raining' expresses the thought that it's not raining. And on Bar-On's neo-expressivism, the second conjunct, 'I think it's raining', expresses the thought that it is raining. So, despite being semantically consistent, Moore's-paradox sentences express incompatible thoughts.

But if such sentences expressed incompatible thoughts, the sentences should carry some sense of literal and explicit contradiction. And, though there is plainly some pragmatic incoherence when we consider saying such things, there is no sense of literal contradiction. And an account is available that, unlike Bar-On's, does not invoke the expressing of incompatible beliefs.

Let's consider a slightly different form of Moore's paradox, though the following explanation applies equally well to other versions. We cannot coherently say 'It's raining but I don't think it is'. Here, the conjunct, 'It's raining', purports to express a thought the speaker has that it's raining, while the conjunct, 'I don't think it's raining', denies that the speaker has that thought. So, as I have argued elsewhere (1995, 2002, 2005: ch. 9), that Moore's-paradox sentence at once purports to express a thought and to deny that that very thought occurs. This explanation invokes the commonsense notion of expressing, on which the speech act and the expressed mental state have the same content, and so sustains the commonsense distinction between expressing a mental state and explicitly reporting that state. And the explanation also sustains the idea that what goes wrong is strictly pragmatic, involving no sense of literal contradiction. We can better explain Moore's paradox without assuming, as Bar-On's neo-expressivist view does, that the remark 'I think it's raining' actually expresses in some way the thought that it's raining.

Bar-On urges (personal communication) that Moore's paradox extends to cases in which one's nonverbal behavior conflicts with one's explicit remarks, for example, fanning oneself while saying that it's not hot. And she urges that the neo-expressivist account of sentences that run afoul of Moore's paradox explains such nonverbal cases as well. But as jarring as such nonverbal cases may be, they do not exemplify Moore's paradox. The absurdity in saying 'It's raining but I think it's not raining' is not present if fan myself while asserting sensibly that it isn't hot. I might well have some compelling reason to fan myself despite my believing that it isn't hot. And saying 'I don't feel hot' while fanning myself is not in any case much like Moore's paradox; I might well know it's hot despite it's not feeling that way to me.

5 Avowals and Reports

Nor need we assume that avowals express the mental states they mention to understand the important, distinctive tie between those states and self-ascriptive

avowals. Let's assume the commonsense view that the speech act 'I think it's raining' reports, but does not also express, one's thought that it's raining, and that the remark, 'It's raining', does express that thought. As Bar-On acknowledges, these two remarks are semantically distinct; they have, as Moore's paradox underscores, different truth conditions. Nevertheless, the two remarks have, with a qualification that does not matter for our purposes, the same conditions of use, that is, the same conditions for the appropriate performance of speech acts that use those sentences. The speech acts are, we can say, performance conditionally equivalent.

The well-known qualification is that the sentence 'I think it's raining' can be used, in contrast to the sentence 'It's raining', to indicate the speaker's hesitation. Bar-On calls attention to this, in a context to which I'll return in §VII. But it's worth mentioning that the opposite also holds; 'I think it's raining', with the right intonation contour, can also indicate dispositive certainty. In any case, degrees of confidence won't be relevant here.

Not only are the speech acts, 'It's raining' and 'I think it's raining', equivalent as regards appropriate use; that performance-conditional equivalence is second nature for us. We use the two interchangeably, so much so that we may not recall even a moment later which of the two we said. And the automatic, second-nature character of the performance-conditional equivalence between the two explains the close tie that avowals have to the states they mention. The avowal, 'I think it's raining', expresses only the thought that one thinks it's raining, not also the thought that it's raining. But its automatic performance-conditional equivalence to the statement, 'It's raining', which does express the thought that it's raining, results in a distinctive, intimate tie between that thought and an avowal that self-ascribes that thought (Rosenthal 2005: chs. 10, 11).

Parallel remarks hold for avowals of purely qualitative states, such as 'I am in pain' and 'It hurts'. We can take those remarks to report, but not to express, one's pain. But, because those self-ascriptions are performance conditionally equivalent to an utterance of 'ouch' and that use equivalence is automatic and second-nature for us, those self-ascribing avowals have a distinctive, intimate tie to the pain itself.

Expressing a mental state is performance conditionally equivalent to self-ascribing that state, and this performance-conditional equivalence is second nature for us. So the intention one has in self-ascribing a state will often be indistinguishable from the intention one might have in simply expressing that state. That fits well with Bar-On's observations about the intentions we have in making avowals. But it does not on that account show that avowals literally express the states they mention.

It will be useful to return to the issue about senses of 'express'. As noted in §III, we need no special notion of expressing to accommodate cases in which one engages in expressive behavior that would typically reveal the presence of a state, but the state isn't there. We can simply say that the behavior, or the person engaging in that behavior, purports to express the state in question. So it's tempting to see the action sense of expressing that Bar-On borrows from Sellars as being simply the notion of purporting to express in the causal sense. Indeed, that makes good on Sellars' idea that the action sense of expressing is indeed a construction from the causal sense. But all this points away from Bar-On's neo-expressivism. In saying something, I purport to express the state that such a speech act ordinarily expresses in the causal sense, and that is so whether or not I am in that state.

Bar-On would reject this characterization of what she has in mind. She sees the behavior a person engages in, whether verbal or not, as action expressing the mental state that behavior ordinarily reveals one to be in. Often I do or say something out of, as we might say, my being in a particular state. I say I think it's raining out of my thought that it's raining, not my thought that I think it's raining; and I say I'm in pain out of my pain, not out of my thought that I'm in pain. Why would I say such things out of my pain and my thought that it's raining if my performing those speech acts didn't express the pain and the thought that it's raining?

The automatic performance-conditional equivalence explains why. Saying 'I think it's raining' is performance conditionally equivalent, in an automatic way, to saying 'It's raining'; so I may say 'I think it's raining' when I might as easily have said just 'it's raining'. I can equally say either out of my thought that it's raining. Similarly, saying 'I am in pain' is appropriate whenever I might say 'ouch', or engage in nonverbal behavior revelatory of pain, and that equivalence is automatic for us. Since I'm roughly equally disposed to say 'I am in pain' or 'ouch' or to engage in such nonverbal behavior whenever I'm in pain, any of those performances will be done out of my being in pain. We need not invoke an additional notion of action expressing to account for all this. The ordinary causal notion, together with the automatic performance-conditional equivalence, suffices.

Bar-On notes (2007, personal communication) that there are circumstances in which the performance-conditional equivalence is not automatic or second nature. I may describe myself as believing that p , for example, not on the strength of my subjectively direct access to that state, but on the basis of some conscious inference, perhaps relying on the word of somebody whom I take to know me very well. And I may in this case not be at all disposed to say that p . So the performance-conditional equivalence it not always automatic. But these cases are relatively unusual, and so do not undermine the foregoing argument, which relies only on its being automatic most of the time. Typically when I'm disposed to say that p , I'm equally disposed to say that I think that p , and conversely.

6 Epistemic Security

What of the unique epistemic security Bar-On takes avowals to have? I'll say just a bit about this, and then close with some remarks, first, about how we learn to make avowals, and then about some connections between avowals and consciousness. Bar-On holds that 'avowals [are] protected from ordinary epistemic assessments (including requests for reasons, challenges to their truth, simple correction, etc.)' (20, 398) and that they are 'strongly presumed to be true' (20, 399). A lot here rides on what being protected amounts to. Behavior seldom gives us enough to challenge or override a person's avowal, though occasionally it does. And that may be all the epistemic security there is. Since avowals are not infallible, challenges are always in place. And, if we had better tools to override avowals, as brain science will in time doubtless deliver, such challenges could be significantly more frequent (for a striking result, see Soon et al. 2008).

Bar-On has urged (personal communication) that the epistemic security she describes is itself a commonsense datum, which we must explain in some way. Neo-expressivism

may then seem the best bet for such an explanation. But whatever commonsense datum obtains is far weaker than Bar-On takes it to be; it is simply that we seldom have reason to override others' avowals, and that when we do have independent evidence, their avowals are usually right. Epistemic security is a theoretical hypothesis about the reason for these commonsense observations. And we should prefer the simplest available explanations of commonsense data, in this case, an explanation that appeals to broadly pragmatic considerations, rather than to theory-laden epistemological claims.

Bar-On builds sincerity into avowing, thereby restricting challenges to cases of misjudgment about what mental state one is in. But that restriction is arguably unwarranted, since even when behavior shows that an avowal isn't true, it may not tell us whether error or dissimulation is responsible. More argument is needed to show that the security Bar-On finds here is not due simply to these kinds of consideration.

Bar-On regards the epistemic security of avowals as 'unparalleled' (20, 399). But that's also arguably doubtful. Past-tense self-ascriptions don't express the states they mention, as Bar-On notes, but my statement that I was in pain two hours ago or that I thought at that time that it was raining are at least as difficult to challenge as their present-tense counterparts.

As Bar-On notes, the privilege of avowals is 'anchored to self-ascribers' (123). But that's just what one would expect from strictly pragmatic considerations. And she notes that this privilege is "inalienable, meaning by this that it is not "easily ... 'detached' from subjects of mental states" (123), which rules out chronic unreliability in subjects' avowals. But a simpler explanation is again available that relies just on the second-nature performance-conditional equivalence I've described to explain why such chronic unreliability does not occur. The psychological pressure to avoid cognitive dissonance doubtless disposes us to think and to say that we think it's raining in pretty much the same circumstances in which we might think and say simply that it's raining. And that by itself will block chronic unreliability in saying 'I think it's raining' and the like.

7 Learning to Avow and Consciousness

Bar-On urges that neo-expressivism can help us understand how we each initially learn to use self-ascribing avowals, such as 'I am in pain' and 'I think that p '. In part, she argues, we learn to use avowals by having them figure as a new form of expressive behavior, replacing, as Wittgenstein (1953: §244) suggests, such natural expressions as wincing or crying out.

But as Bar-On also notes, that model will not do in general, since there are no natural expressions for most mental states. So she urges that we learn, as young children, to use avowals from having others apply mental descriptions to us (288). In particular, we learn to say 'I think that p ' when it seems to us that p is so, but we aren't quite sure (294). The hesitation often suggested when we say 'I think' figures in our initially learning, as children, to use such avowals.

As I noted earlier, in §V, hesitation is by no means always suggested when we say that we think something. But that aside, there is a difficulty with this part of Bar-

On's suggestion. As Josef Perner and colleagues (1987; Wimmer and Perner 1983) have shown, children initially take thinking and believing always to correspond to what is actually the case, learning only between ages 4 and 6 that a person's thoughts and beliefs sometimes depart from reality. Children do not initially learn to use 'I think' to convey hesitation.

Still, Bar-On's general idea is doubtless sound, that young children initially learn to use avowals from their elders' applying mental terms to them. But that, by itself, is neutral as between her neo-expressivist theory and its descriptivist alternatives. When adults ascribe mental states to a child, the child may well preserve the purely descriptive nature of the adults' ascriptions in adapting them to the first-person, just as children do with ascriptions of bodily and perceptual conditions.

The Perner findings do suggest that young children initially equate saying that p and saying that one thinks that p . What happens, then, when children do learn, between ages 4 and 6, how the truth conditions for the two diverge? One could speculate, as Bar-On does (personal communication), that their use of 'I think that p ' continues to express the thought that p , as against coming to express only the thought that one thinks that p . But without some empirical reason to think that, this speculation simply reflects, rather than supports, the neo-expressivist theory.

Bar-On sees descriptivist theories as requiring an indefensible view about how we are able to say of ourselves that we are in one or another mental state. By contrast, she argues, neo-expressivism provides an acceptable answer to that question. Descriptivism, she urges, must assume some quasi-perceptual, recognitional ability to tell what mental state one is in. Self-ascriptive avowals would then simply express the descriptive content of that quasi-perceptual recognition that one is in the relevant state. Bar-On argues forcefully, and to my mind convincingly, that no such quasi-perceptual, recognitional capacity exists. And she stresses that neo-expressivism obviates the need to appeal to any such capacity, since one's ability to avow being in a particular state is on that view simply of one's ability to express that state.

Some descriptivists do appeal to a recognitional capacity. Indeed, on a theory of consciousness that invokes higher-order perception, such as William G. Lycan's, it's natural to appeal to some quasi-perceptual recognition of conscious states, as Lycan (2004: 109) has recently done. But that's a signal difference between such inner-sense theories and my own HOT theory. Though both theories are purely descriptivist, I join Bar-On in rejecting any appeal to a recognitional capacity (e.g., Rosenthal 2004: §3).

Instead, the HOT theory relies here on the performance-conditional equivalences I mentioned earlier. We come to learn that it's appropriate to say 'I am in pain' in just those circumstances in which one might say 'ouch', and to say that one thinks that p in just those circumstances in which one might simply say that p (Rosenthal 2005: ch. 10; 2008). But, as Moore's paradox shows, two speech acts can have the same conditions for appropriate use without expressing the same mental state. If the speech acts, 'It's raining' and 'I think it's raining', did express the same state, Moore's paradox would be contradictory, rather than just pragmatically incoherent. The automatic performance-conditional equivalences do the work of Bar-On's assumption that avowals actually express the states they mention, but without the disadvantages of that controversial claim.

It's worth noting that the performance-conditional equivalence between saying 'ouch' and 'I am in pain' and between saying that p and saying that one thinks that p

is independent of the Perner findings. The Perner results concern when children come to recognize that saying that p and saying that somebody thinks that p have distinct truth conditions, not whether they recognize that two such speech acts are performance conditionally equivalent.

If each avowal did, as Bar-On insists, express in different ways both the state it mentions and a thought or judgment that one is in that state, that would explain this automatic performance-conditional equivalence. Saying that p would then be automatically equivalent in performance conditions to saying that one thinks that p . But there are, independent of that neo-expressivist claim, no circumstances in which it's appropriate to perform one of those two speech acts but not the other. Indeed, if that were not so, there would be nothing absurd about Moore's paradox. So explaining the performance-conditional equivalence does not require neo-expressivist assumptions.

Nor are neo-expressivist assumptions needed to explain the automatic character of that performance-conditional equivalence. On the compelling picture developed by Sellars (1956), we come to describe ourselves and others as being in intentional states by initially positing such states as the causes of speech acts and rational nonverbal behavior. This account arguably applies not just to the way humans originally came to describe themselves as being in intentional states, but to individual development as well (Sellars 1975: §II). Seeing intentional states as causes of speech acts and dispositions to speak underwrites ascribing such states whenever the corresponding speech acts occur or are disposed to occur. And that will induce a disposition to ascribe an intentional state to oneself whenever one is disposed to perform the corresponding speech act. This in turn will result in the performance-conditional equivalence's being relatively automatic, as well as generating accurate HOTs about one's intentional states (Rosenthal 2005: ch. 10, §5; 2008: §5).

Bar-On expresses doubts about my HOT theory, urging that it's unintuitive to hold that the states of creatures are conscious only if those creatures can make higher-order judgments about what states they are currently in (354, n. 12). Calling the HOTs judgments certainly helps make that seem unintuitive, since it makes HOTs seem cognitively elaborate and epistemically sophisticated. I've argued that only a rather minimal notion of thought is needed here, and that the HOTs themselves need not be, and seldom are, conscious.

But it's worth mentioning an argument (Rosenthal 2005: ch. 2) for HOTs that relies on premises that seem hard to contest and that pertain to the expressing and reporting of mental states. For creatures, like adult humans, who can report being in mental states, it's standard to hold that a state is conscious if, but only if, it's sincerely reportable. Sincere reports, that is, avowals, occur only when the states are conscious. If somebody is in some mental state but sincerely denies being in it, that state does not count as conscious.

But a report of being in a state expresses one's thought that one is in that state. So the ability to report being in a state corresponds to the ability to express a thought that one is in that state—that is, a HOT. So mental states arguably are conscious if, but only if, they're accompanied by HOTs that enable one to issue such sincere reports. And having used reportability to fix the reference of the term 'conscious state' for creatures, like us, who can say what mental states they are in, we can then extend the HOT account to creatures who cannot.

Bar-On says little about mental states that aren't conscious or about what distinguishes them from states that are conscious. That's understandable given her focus on verbally expressed states. Verbally expressed intentional states are, with a minor exception, always conscious, unlike intentional states that are expressed only nonverbally (Rosenthal 2005: ch. 10). Still, mental states do occur without being conscious. And if the HOT theory is wrong, we need some other way to explain the reportability test.

On Bar-On's neo-expressivism, remarks like 'I am in pain' and 'I think it's raining' both describe oneself as being in pain and as thinking that it's raining and literally express those very states. Her challenging, admirable book is doubtless the most thoroughgoing and well-developed defense of that view. That view would if sustained show how such remarks share features of both reports and expressions of the very same mental states. But given the difficulties that face Bar-On's neo-expressivism, we should regard such remarks as simply describing or reporting the states they are ostensibly about, and not also as expressing them.

References

- Bar-On, D. (2004). *Speaking my mind: expression and self-knowledge*. Oxford: Clarendon Press.
- Bar-On, D. (2007). *Reply to Matthew Boyle and David Rosenthal*. Presented at Author-Meets-Critics session, American Philosophical Association, Eastern Division, December 2007.
- Frege, G. (1977). Thoughts. In P. T. Geach & R. H. Stoothoff (Eds.), *Logical investigations* (pp. 1–30). New Haven: Yale University Press.
- Gasking, D. (1962). Avowals. In R. J. Butler (Ed.), *Analytical philosophy I* (pp. 154–169). Oxford: Basil Blackwell.
- Lycan, W. G. (2004). The superiority of hop to hot. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness* (pp. 93–113). Amsterdam and Philadelphia: John Benjamins Publishers.
- Moore, G. E. (1942). A reply to my critics. In P. A. Schilpp (Ed.), *The philosophy of G. E. Moore* (pp. 533–677). New York: Tudor Publishing Company.
- Moore, G. E. (1944). Russell's theory of descriptions. In P. A. Schilpp (Ed.), *The philosophy of Bertrand Russell* (pp. 175–226). New York: Tudor Publishing Company.
- Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year olds difficulty with false belief: the case for a conceptual deficit. *British Journal of Developmental Psychology*, 5(2), 125–137.
- Putnam, H. (1968). Brains and behaviour. In R. J. Butler (Ed.), *Analytical philosophy, vol. II* (pp. 1–19). Oxford: Basil Blackwell.
- Rosenthal, D. M. (1995). Self-knowledge and Moore's paradox. *Philosophical Studies*, 77, 2–3. 1994 Pacific Division APA Special Issue (March): 195–209.
- Rosenthal, D. M. (2002). Moore's paradox and Crimmins's case. *Analysis*, 62(2), 167–171.
- Rosenthal, D. M. (2004). Varieties of higher-order theory. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness* (pp. 17–44). Amsterdam and Philadelphia: John Benjamins Publishers.
- Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Clarendon Press.
- Rosenthal, D. M. (2008). Consciousness and its function. *Neuropsychologia*, 46(3), 829–840.
- Sellars, W. (1956). Empiricism and the philosophy of mind. In W. Sellars (Ed.), *Science, perception and reality*. London: Routledge & Regan Paul Ltd., 1963.
- Sellars, W. (1969). Language as thought and as communication. *Philosophy and Phenomenological Research*, 29(4), 506–527.
- Sellars, W. (1964). Notes on intentionality. *The Journal of Philosophy*, LXI(21), 655–665.
- Sellars, W. (1968). *Science and metaphysics*. London: Routledge & Kegan Paul, and Atascadero: Ridgeview Publishing Co., 1992.
- Sellars, W. (1975). The structure of knowledge, lecture II: minds. In H.-N. Castañeda (Ed.), *Action, knowledge, and reality: critical studies in honor of Wilfrid Sellars* (pp. 316–331). Indianapolis: Bobbs-Merrill.

- Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*(5), 543–545.
- Strawson, P. F. (1959). *Individuals: an essay in descriptive metaphysics*. London: Methuen & Co., Ltd., ch. 3.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.
- Wittgenstein, L. (1953). *Philosophical investigations*, ed. G. E. M. Anscombe and R. Rhees, tr. G. E. M. Anscombe. Oxford: Basil Blackwell.

What follows derives, slightly revised, from 'The Mind and Its Expression,' to appear in a volume in honor of Jay Rosenberg, edited by Eric M. Rubenstein and James R. O'Shea. I am grateful to the editors for permission to use this material. My discussion here also expands on comments on Bar-On's (2004) at an Author-Meets-Critics session, American Philosophical Association, Eastern Division, December 2007.

Copyright of Acta Analytica is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.