

# Replies Part II: Knowledge of Our Own Minds and Meanings

*Bar-On and Smith*

## The Problem of Self-knowledge

Although some philosophers may have preferred to think otherwise,<sup>1</sup> it is not philosophy but part of the ordinary folk notion of the mental that each of us stands in a special cognitive relationship, denied to others, to our own thoughts, hopes, imaginings, and sensations—that our mental lives are directly available to us and only indirectly available to others, that “You cannot really know what another is thinking”, whereas of one’s own thoughts one cannot but be aware. This notion is a *premise* for the other minds problem. It is captioned in much early analytical philosophy of mind by the phrases, “privileged access”, and “first-person authority”. The conception of the problem of self-knowledge addressed in the work of mine<sup>2</sup> that Dorit Bar-On’s and Barry Smith’s chapters respond to is essentially that of achieving a satisfactory perspective on what is right about the idea of “privileged access”, a perspective to head off the beckoning slide into other minds scepticism, and the problems brought to the fore by Wittgenstein’s discussion of private language, while at the same time finding a place for (perhaps disinfected versions of) the seemingly undeniable asymmetries that motivate the folk notion.

It was, I think, Ryle who introduced the use of the term, ‘avowal’, in this context. Since *The Concept of Mind*, the major tendency of philosophical discussion of psychological self-knowledge<sup>3</sup> has been to focus on its linguistic expression and on the asymmetries as reflected in characteristics of the competent use of avowals, in contrast to the competent ascription of mental states to others. If I say that you are in pain, or that you are expecting a postal order in the mail, or anxious about the economic condition of English Premier League football, I make a claim for which it is appropriate to ask for my reasons. But if I avow any of those claims of myself, the request for reasons, it is commonly said, “makes

<sup>1</sup> I am thinking of John McDowell, but perhaps unfairly.

<sup>2</sup> There isn’t a lot: essentially just the Whitehead Lectures, and the substantially overlapping “Self-knowledge: The Wittgensteinian Legacy”. But I have thought, and continue to think about the issues a lot more than this rather meagre return indicates.

<sup>3</sup> In general, I’ll just say “self-knowledge”.

no sense". Again, if I make any of those claims about someone else, it is straightforward to specify a context in which you might, without doubting my sincerity, reasonably wonder whether what I was saying was true. But when the corresponding claims are made of myself, it is at least not straightforward, and may in context be impossible, to envisage circumstances in which, without doubting my sincerity, you might reasonably wonder about the truth of what I have said. If, finally, I have a headache, or if I am concerned about the financial predicament of English Premier League football, you would normally expect me to be in a position to say so, and, *ceteris paribus*, willing to say so. But there would be no normal expectation that *you* would be in position, or willing, to say so on my behalf.

These are, at first exposure, points about our linguistic practice. No doubt they could stand refinement and nuance to accommodate details of context and variations the kinds of mental state concerned, but there seems no prospect that they should simply prove *illusory* on closer examination. At this point though, we confront a fork. What comes first here: the linguistic practice, or the thoughts of the thinker manifested in that practice? The problem of self-knowledge will look different depending on how one chooses. On the first option, we will tend to think of matters—very much as Ryle and Wittgenstein did, and as I think Bar-On does too—as centred upon explaining avowals. The problem will be viewed as that of accounting for the distinctive aspects of the “grammar” of avowals, in contrast to the competent other-ascription of mental states; and a range of candidate explanations will become salient, for instance those falling under the broad rubric of expressivism, which will seem point-missing on the second option. On the latter option, by contrast, the relevant features of avowals will, from the outset, be seen as reflections of certain special aspects of the epistemic character of the self-directed thoughts they express. And it will seem overwhelmingly natural to suppose just what the folk notion does suppose: that selves characteristically know of the states that give rise to avowal in a way that involves no inference or independent reasons, and which is characteristically very secure; and that the states of the relevant kind are typically salient to their subjects. And the problem will then be to account for these apparent epistemic advantages in a way that is proof against a slide into Cartesian privacy and its associated nemeses.

The triptych of *immediacy*, *authority*, and *transparency* that I have used to configure the explananda of the problem allows of interpretations to accord with either alternative: language first, or thought first. Let us pick up the old term ‘privileged access’ as a name for the triptych on its epistemic interpretation. It’s worth pausing a moment over the notion of immediacy as an ingredient of privileged access. The distinction between inferential and non-inferential judgment is itself in need of clarification but immediacy should here not be equated with non-inferentiality in any case.<sup>4</sup> If, for example, we take perceptual

<sup>4</sup> McDowell, in “Response to Crispin Wright” (1998), takes it that this distinction rapidly undercuts any diagnosis of the appeal, for folk philosophy, of a Cartesian, observational model as an apparent explanation of privileged access. I reply that the distinction is a difficult one, and it is only too plausible that ordinary thought would miss it.

judgment as a paradigm of the non-inferential, there is still scope for views which hold that perceptual judgments nevertheless have a basis in reasons: reasons provided not by other beliefs but by states of perceptual awareness (or perceptual seeming). Such a state—its non-doxastically seeming to one that P—is commonly regarded as rationalizing the belief that P even in cases where it does not consist in a perceptual awareness that P. But for psychological states, in contrast, there is in general no plausible candidate for such a mediating, rationalizing state. In the case of sensation, for example, there is no evident distinction to be drawn between a non-doxastic seeming that, for example, one has an itch between one's toes and one's actually having the itch. And in the case of, say, the belief that English Premier League football is financially unsound, there seems no sense at all to be attached to the idea of a *non-doxastic* seeming that one has that belief—though one may of course believe that one believes it.

### Constitutive Views

My exploration of so-called Constitutive views was fashioned as a strategy for evading the Language First or Thought First dilemma. In effect, the idea was to point out a third, deflationary option: to side with Thought First, but without the invocation of privileged access. On such a view we will essay to regard those of a subject's self-directed thoughts that potentially issue in avowal as indeed having properties that underwrite the characteristic features of avowals, but not the properties associated with privileged access, indeed not properties of epistemic provenance at all. The choice of the term, "constitutive", was perhaps not entirely felicitous. At any rate, it has encouraged unnecessary criticisms. The idea was not at all that a toothache, say, might be in part constituted by the judgment that one has a toothache, in a metaphysical sense like that in which the identity of the singletons of Cicero and Tully might be conceived as in part constituted by the identity of Cicero and Tully—as if the fact of the toothache itself had a judgment as a component. The suggestion was, rather, that the distribution of truth-values among propositions concerning a subject's mental states is, a priori and necessarily, constrained by what she herself takes to be the truth about her mental states, and that this is a conceptually basic or *primitive* point, rather than a consequence of some independently accountable aspect of epistemic advantage which selves enjoy in relation to their own states of mind. The master thought of Cartesianism, epitomized by the metaphor of the inner theatre and the very etymology of the term, "introspection", is that there is indeed such an advantage. The project of any Constitutive view is to do without that, or anything like it, while yet avoiding falling back on the Language First alternative.

I think it is an open question what is the most robust formulation of a view of this type. In my own earlier work I experimented with two.<sup>5</sup> One was to invoke the model

<sup>5</sup> A third form of Constitutivism is of course at the heart of Akeel Bilgrami's 2006. I have no space to compare and contrast the details here.

of response- or judgment-dependence explored in *Truth and Objectivity* and elsewhere. According to that model, certain subject-matters—perhaps colour and secondary qualities generally, perhaps certain kinds of value—have the feature that the extensions of their signature concepts are determined, at least in part, by suitably constrained sensory, affective, or doxastic responses of ours; in the case of judgment-dependence, by the very (suitably constrained) judgments of ours about what the concepts in question apply to. “Suitably constrained” means that the responses in question only count as appropriately extension-determining if they are elicited under certain independently specifiable optimal, or normal conditions. Thus, as a putatively prototypical example, the surface of an object observed under optimal lighting conditions by an attentive, normally visually functioning individual is red if and only if he is thereby affected with the appropriate distinctive visual experience. To regard the extension of *red* as response-dependent is to regard (a suitable specification of) this rubric as an a priori conceptual necessity, and one whose necessity is owing, moreover, not to the fact that redness infallibly *causes* experiences of the relevant kind under the conditions in question, but to the fact that there is *nothing more* to an object’s being red, under the conditions in question, than its so appearing to a normally sighted subject—the well-known form of contrast first outlined in Plato’s *Euthyphro*.

This broad idea can naturally be—and has been—implemented by a number of specific templates differing in detail. To bring it to bear to deflate the pressure towards invoking privileged access, a template involving some form of biconditional dependence will clearly be required—so that there is a chance of mimicking both transparency and authority—and the relevant response will have to be that of judgment, since the connection that we are trying to safeguard is between the psychological facts and the subject’s impression of them. Because the required connection has to be highly reliable, the relevant optimal conditions will have to be ones that normally obtain, or can very easily be brought about. But we do not want complete reliability, since failures in the direction of transparency may occur in, for example, conceptually limited subjects, and failures in the direction of authority may occur as a result of phenomena of priming, self-deception, and so on.

That is one constitutivist model. But there is a problem with any such proposal, first pointed out by Paul Boghossian.<sup>6</sup> Simply: in this case the relevant responses have, as noted, to consist in certain kinds of judgment, and judgment is of course itself an intentional psychological phenomenon. It is important to be clear why exactly this is a problem, given that on some models of response-dependence,<sup>7</sup> subjects’ responses under optimal conditions are in any case thought of as determining only *part* of the extension of a relevant concept. Such partiality is not the concern. On such models, there is still scope for this partial determination to be from *without*, as it were, by facts of

<sup>6</sup> See Boghossian 1989, pp. 544ff. Also the present volume at p. 34.

<sup>7</sup> Those that work with what in *Truth and Objectivity* I termed *Provisional Equations* (or ‘provisoed biconditionals’).

a different kind. But on any judgment-dependence proposal about ordinary psychology of the broad stripe presently considered, the responses—the determining facts—are, qua judgments, further examples of facts of *the very kind* supposedly being determined. The proposal is committed to the thought that the distribution of truth-values among ascriptions of (first-order) judgment to a thinker is determined, in part, by his second-order judgments. What determines the distribution of truth-values among ascriptions to her of these second-order judgments? If we reapply the model, it looks as though we now need to fall back on her *third-order* judgments . . . and the beckoning regress, and associated spectre of ungroundedness, threatens to undermine the very dependence postulated by the Constitutive view. But if the model is not to be reapplied, then the theorist is building in an unexplained exception to the judgment-dependence proposal, and now faces embarrassing questions about what fixes the distribution of truth-values among ascriptions of the relevant second-order judgments, and about their first-personal epistemology.

It was these considerations that moved me towards a different implementation of the basic thought of the Constitutive view, one that promises to finesse Boghossian's observation. Donald Davidson is famous for the *interpretationist* proposal that the facts about a subject's mental states do not outrun whatever would be opined by an ideal radical interpreter, operating under conditions of maximum information. This kind of proposal, of course, has problems of its own but—assuming that they are not in the end lethal—it looks to allow for a ready extension, not apparently envisaged by Davidson himself, to address the present concerns. Simply: it can be taken to be a *constitutive principle* of best psychological interpretation that the interpreter must maximally respect the express self-conception of the interpretee (authority) and must minimize the extent to which unacknowledged mental states are ascribed to the interpretee (transparency), whilst otherwise making the best possible sense of what she says and does. In effect, rather than being charged simply with making the most satisfactory overall sense of the subject's linguistic and non-linguistic behavior, the interpreter is required, in doing so, to keep to a minimum the extent to which the mental states ascribed conflict with, or exceed, those the subject herself is willing to acknowledge. Perhaps better: compliance with this constraint is taken as a necessary condition of “making the best possible sense” of what the subject says and does. The constraint is to operate holistically, with data of apparent acknowledgement themselves regarded as defeasible and open to interpretation as part of the same exercise. That is enough to parry the immediate concern about ungroundedness that confronted the judgment-dependence model: the determination is achieved at the level of whole end-product, rather than step-by-step by individual judgments.<sup>8</sup> To be sure, there are still concerns about underdetermination, and about

<sup>8</sup> To parry, but not perhaps to finally assuage: for have we not merely replaced the dependence of the subject's states on her second-order attitudes with a dependence of them on the attitudes of a third party—the hypothetical best interpreter?

the inexplicit character of the methodology of best interpretation, but they were there from the start.

What is the relation of any of this to Wittgenstein's thought? There is of course no basis in his texts for ascribing to him any explicit form of response-dependence or interpretationist proposal. But a contention that I think he did hold—or which anyway falls out very quickly from an application of the metaphilosophy of the *Philosophical Investigations* to the present issues—is, though similarly deflationary in spirit, in tension with the Thought First aspect of Constitutivism in any case. It is the contention that our difficulties in this area result from the misguided quest for *philosophical explanation* of what are in fact basic features of the “language game” of ascription of psychological states to oneself and others. In Wright 1998 I called the position that develops out of this idea the *Default View*. The Default View is a Language First view. It holds that the authority, for example, that attaches to a subject's self-conception of her own mental states is an epiphenomenon of the authority invested, by the rules of the language game, in what she has, sincerely and comprehendingly, to say about her own mental states. This is not expressivism, though it is consistent with an expressivist embellishment of certain types of avowals—which, to the gratification of commentators such as Bar-On, Wittgenstein himself here and there famously seems to go in for.<sup>9</sup> On the Default View, the aspects of linguistic practice that (misguidedly) motivate the notion of privileged access are not to be explained in terms of other aspects of their semantic role, for example by their being “expressions”. They are *not to be explained at all*. It is rather that by being primitively endowed with authority, they naturally acquire the informational significance possessed by the canonical behavioral expressions of a subject's mental states.

So let's take an overview of the landscape of the ways of avoiding privileged access just distinguished. Neither version of the Constitutive view—judgment-dependence or interpretationist—is directly concerned with the phenomenon of avowal. On either proposal, a *metaphysical* thesis is being advanced about the nature of the psychological: that what is true about a subject's psychological states is constitutively conditioned by what she herself, in normal circumstances, takes to be the truth about them. These proposals are no more concerned with our characteristic ways of speaking than are the contentions that colour is response-dependent, or that linguistic meaning is interpretation-dependent. However, either contention will entail, and in that sense explain, the special status accorded to a subject's avowals. Expressivism also gives an explanation of that—but one which derives not from the nature of the subject-matter, but from the

<sup>9</sup> The most famous passage is of course *Philosophical Investigations* §244: “. . . how does a human being learn the meaning of the names of sensations?—of the word ‘pain’ for example? Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behavior.” But this is as close as Wittgenstein gets to an explicit expressivism, and there is nothing in the passage or its surroundings to exclude the “new pain-behavior” consisting in a practice of assertion.

alleged distinctive *semantic* role of avowals. But on the Default View, as a special case of Wittgenstein's exhortation of the urge to philosophical explanation, the distinctive characteristics of avowals really are grammatically primitive, as it were; there is no explaining them and if any plausibility attaches to response-dependence, or interpretationist, or expressivist views, it is because they offer prismatic, somewhat distorted reflections of this primitiveness.

Now, it is not completely clear whether, when Bar-On is critical of what she understands by the Default View, and when Smith is critical of what he terms my "deflationary" approach, either has in mind the distinctions just noted. Smith's principal concern is with self-knowledge of meanings, but his complaint against my discussions seems to be a generic concern, applicable to all the views above: it is that such views do not really address the question of self-*knowledge*. If this complaint is addressed against Constitutive views as characterized above, then it is misguided. There are conceptions of knowledge to hand—those which, for example, centralize reliability, or safety—which will readily allow those of a subject's opinions which meet appropriate conditions, determined by a conception of their subject-matter as judgment-dependent, or as interpretation-dependent in the kind of fashion proposed, to rank as knowledgeable. It is of course true that knowledge, so achieved, will not be a matter of *keeping track*. But of course the whole point of such views is to repudiate the notion that self-knowledge is, in the basic avowable case, a matter of such keeping track. The objection is yet more tendentious if directed against the Default View, a major part of whose point is, in effect, precisely to repudiate the idea of self-knowledge in the first place; on this view, the special place occupied by avowals in the language game of ascription of psychological states is exactly *not* a reflection of the knowledgeability of their authors.<sup>10</sup>

For her part, Bar-On's objections to the Default View are various. Sometimes it seems that it is some kind of constitutive view that she has in mind, as when she complains that it is not clear what exactly is being offered by way of analytical reconstruction of the truth-conditions of ascriptions of mental states<sup>11</sup> The objection that seems to weigh most forcefully with her, however, is different and, I think, important. Not all self-ascriptions of mental states exhibit the characteristic marks of avowals. There are ascriptions of mental states to oneself made at times other than that of the occurrence of the state concerned; and there are what Bar-On calls "reportive" such ascriptions based on reasons, interpretation, and evidence. No account, she charges, which simply privileges those of a subject's opinions about herself which are formed under good but normal conditions,

<sup>10</sup> Thus *Philosophical Investigations* §246: "It can't be said of me at all (except perhaps as a joke) that I *know* that I am in pain. What is it supposed to mean—except perhaps that I *am* in pain?"

<sup>11</sup> This volume, p. 177.

or privileges her self-conception in the interpretation of her by others, or simply privileges what she has to say about herself, can provide the resources to explain this distinction.

Is this a compelling objection? On the face of it, it would seem that it cannot be: that if we had an adequate characterization of the data that set the problem—that is, of the range of cases, whether in speech or in thought, that apparently manifest immediacy, authority, and transparency—then this characterization, whatever it is, could simply be built into the appropriate formulation of the relevant template for judgment-dependence, or the appropriate constraint on interpretation, or the appropriate specification of the rules of the language game of ascription of psychological states. Only characterize the problem properly, it might be thought, and any view of the kinds I have been distinguishing will have been given the resources to respond to Bar-On's concern.

This looks like the right direction for a reply to Bar-On's objection to take. But a caveat is called for. What exactly *is* an avowal? Is there a complete and correct characterization of this class of utterances? The interest of the judgment-dependence and interpretationist proposals is in a certain class of opinions, rather than their linguistic vehicles. But it looks as though it would be difficult to say which exactly the relevant opinions are—which are the opinions which, when formed under suitable conditions, are to be conceived as partially determining the facts of a subject's ordinary psychology, or as fixing those aspects of her self-conception with which an optimal interpretation is constrained to accord—other than by saying that they are those opinions whose expressions constitute a proper avowal? But then we have to reckon with the following possibility: that a serious attempt to give a comprehensive account of which are the first-personal self-ascriptions that exhibit the characteristic features must, if it is to do better than a mere illustrative list, advert to properties of these utterances which favor an account that is at odds with the broadly deflationary kinds of proposals that I have been reviewing. It might be charged, for example, by someone with vestigial Cartesian sympathies that there is ultimately no characterizing the class of first-personal self-ascriptions that may be properly avowed other than in epistemological terms: an utterance of this kind can be properly avowed just when the subject stands in the appropriate distinctively first-personal epistemic relation to the state of affairs it depicts. At the other extreme, and more congenially to Bar-On, it might turn out that there is no comprehensively characterizing the relevant class save by reference to the notion of an utterance that *expresses*, according to a preferred account of that notion, the relevant mental state; that the characteristic marks of avowals are grounded in, and their extent determined by, the illocutionary act of expression. Accordingly, it does seem fair to say that a proponent of either form of deflationary view does have a pressing obligation to show that a characterization is possible of the target range of utterances, or opinions, which avoids letting in the opposition by the back door, as it were. I would accept the criticism that, at the time of writing, this obligation has not really been discharged.

## Expressivism

There has been no more determined or systematic a supporter of expressivism as a response to the problems of self-knowledge than Bar-On. Her contribution to this volume outlines key features of her major elaboration and defence of the view offered in her 2004 and elsewhere. The key move is one that I myself argued for in Wright 1998 and I agree with her that it is a vital part of any viable expressivism in this region. This is to jettison altogether the idea, originally the cardinal point of, for example, ethical expressivism, of any tension between the allegedly expressive role of avowals and the acknowledgement of a domain of ordinary psychological truth and falsity. The thesis has to be not that there is no such thing as genuinely true or false psychological assertion, *including* assertion about one's own case, but that those aspects of our discourse which give the impression of being an especially authoritative genre of such assertion are actually not assertoric at all, serving rather to, in a relevant technical sense, give expression to the beliefs, or feelings, that, *prima facie*, they seem to state. In those parts of her chapter, accordingly, where Bar-On is arguing that a competitive expressivism about avowals must distance itself from the non-factualism of traditional expressivist views of, for example, moral value, she is, at least as far as this reader is concerned, preaching to the converted.

That agreed, I am however doubtful about Bar-On's suggestion that expressivism about avowals, so liberated from any non-factualist implication, can then comfortably coexist with a minimalist conception of truth and truth-aptitude for ordinary psychological discourse in general. In Wright 2002 I argued in some detail that minimalism about ordinary intentional psychology is actually a dubiously stable position dialectically, since it appears to entail a corresponding minimalism about the very distinction between merely minimally truth-apt discourses and those that discharge a more robustly representational role, and that it is the hallmark of the former that they fail to exert cognitive command, that is, that there is no *a priori* guarantee the disputes within them need involve anything worth regarding as a cognitive shortcoming. The argument I gave is quite complex, and is no doubt open to challenge. But unless it is addressed, the bottom line threatens that to regard ordinary psychological discourse as, though truth-apt, merely minimally so is a commitment to regarding that classification of it as rationally unforced. That is not a contradiction, but it does raise the question, of one who takes this view, why exactly they are doing so—something close to, though not quite, a kind of Moorean paradox.

Much better, then, if expressivism in the present context is yoked to a realist view of ordinary psychological claims—as indeed, in Bar-On's own exposition, it is. I would not still wish, without attempting to respond in detail to the development and defence in Bar-On 2004, to stick to my sometime description of this type of view as a “dead duck”. But I continue to have a number of reservations that go to the basic structure of

the view and are independent of its detail, and which strike me as serious enough to make it unlikely that any version of it can be fit for purpose. Bar-On herself has fashioned replies to some of these objections, which she refers to but does not enlarge on in her present chapter.<sup>12</sup> So I'll provide merely the briefest outline of them before moving to a new point.

There is much to be said about what notion of expression best serves the expressivist's theoretical purposes, and Bar-On's discussions contribute considerably to the clarification of this issue. But for the purposes of any expressivist realism equipped to dissolve the problem of self-knowledge in the kind of way anticipated, the essential point has to be the following. Selves must turn out to stand in no form of epistemically superior relationship to their own mental states that would ground the distinctive features—those suggestive of privileged access—of avowals. *Those* features have to be explained by showing how they flow from the expressive function of avowals. So *expression* has to be understood in a way that makes that possible. The idea of any distinctively privileged first-person epistemology of the psychological has to be undercut; *that is the whole point of the view*. This, of course, doesn't debar one who regards avowals as having a straightforwardly assertoric role from acknowledging that they also have an expressive function—just as a realist about moral value is not debarred from allowing, if it is true, that moral assertions characteristically also serve to express feelings of moral approval and disapproval. But that acknowledgement does not amount to expressivism in the sense here of interest. The crucial move must be to utilize the expressive character of avowals to obviate any impression that their distinctive features stem from some kind of recognized first-personal epistemic superiority—in effect, to show the problem of self-knowledge, conceived as I have outlined it, is an *illusion*.

I am afraid I remain utterly sceptical that this can be done. In earlier work, besides emphasizing that the distinctive attributes of authority, immediacy, and transparency seem to belong as much to the unvoiced *impressions* a subject has of her own mental states as to her expressions of them, I noted how the idea that all a thinker might have to go on, as far as her strictly epistemic situation is concerned, in her assessment of her own mental states are the public expressions that are available to anyone, leads to absurdity in certain kinds of scenario, where the outward behavior manifesting a certain psychological state—perhaps acute social embarrassment—is actually more salient to onlookers than to the subject who is trying to conceal it.<sup>13</sup> One, as it seems to me, rather desperate response to the first of these considerations is to attempt to regard token thoughts concerning one's own mental state as themselves a kind of expression—a kind of interior avowal, if you will.<sup>14</sup> But such a move, it should be evident, even if allowed, makes no progress at all with the second problem. For one thing, it seems that my awareness of a toothache is hardly likely to be compromised if

<sup>12</sup> This volume, p. 176, fn. 40.

<sup>13</sup> Thus the torture victim case in Wright 1998, p. 37.

<sup>14</sup> See this volume at p. 185 and Bar-On 2004, p. 23.

I fail to token any relevant thought—say, “This tooth hurts like Hell!”—and thus simply don’t express it at all, even inwardly. And even if I do happen to ‘inwardly express’ my pain by such a thought, and so have that additional datum, that will hardly restore my epistemic situation with respect to the pain to one of superiority over onlookers unless I am aware of *the thought*—and aware of it in what appears to be a distinctively first-personal way, since the case is, by hypothesis, one where I don’t articulate it publicly.

Let me, though, move to the advertised new point. Behavior that is prototypically expressive of a mental state, in the kind of way Wittgenstein was thinking of—grimacing and hopping about while clutching a stubbed toe, scratching at a mosquito bite, or a child’s wowing after it has dropped its ice-cream on the path—is involuntary. More sophisticated kinds of expressive behavior—for instance, swearing, or hurling down the newspaper when reading the football results, or grabbing the bars of one’s cell and making as if to shake them loose—are often, by contrast, intentional: subject to voluntary control and in some sense selected. However such behavior, although intentional, does not, insofar as it is expressive of the reactions or emotions portrayed, amount to *action supported by a structure of reasons*—beliefs, desires, and intentions—that coordinate into a practical syllogism. In that sense, such actions are not properly classified as *rational* actions, though they are rationally intelligible acts. For convenience, let’s term the three categories *involuntary*, *intentional*, and *purposive* respectively. Now in principle, a vocalization with a truth-evaluable content can belong to any of the categories. Wittgenstein’s idea is at least coherent: there is no reason a priori why an episode of linguistic behavior, of a kind acquired as a replacement for the natural expressions of a pain, would have to be intentional, let alone purposive. The point that needs to be noted, however, is that the *mode of evidential significance* possessed by such a performance must vary as a function of which category it is taken to belong to. It is of course possible to simulate pain behavior, or voluntarily “let it out”, or do so as part of an attempt at communication with a foreign doctor. But part of the irresistible evidential force of pain behavior, at its most convincing, is precisely the consideration that it is conceived as involuntary, as a result of the pain’s surging to the surface, as it were, so the question of insincerity does not arise. Avowals, however, cannot in general acquire their evidential force on that model, even when the mental state in question does have natural untutored forms of behavioral expression. For we can take it as a datum that avowals will only exceptionally be involuntary. Usually they will be at least intentional, even if not purposive. But most often they will be regular speech acts, underwritten by a practical syllogistic structure of reasons.

What follows? Granting that much is still consistent, no doubt, with there being some theoretically pointful notion of expression such that a deliberate, purposive avowal, underwritten by a structure of practical reasons, may still count as an *expression* of the relevant mental state—that state, namely, that satisfies the truth-conditional content which, Bar-On’s kind of expressivism grants, is associated with the avowal. What seems clear, though, is that this notion of expression, if such there be, won’t be

able to deliver the intended upshot that the impression of epistemic superiority conveyed by misassimilating an avowal to an assertoric report should turn out as illusory. To appreciate this we have merely to notice that among the subject's set of reasons for a sincere purposive avowal, confidently made, will have to be an awareness of the relevant psychological state. Your General Practitioner asks you, "Point out where it hurts". Clutching your lower right abdomen, you say, "It hurts here, especially when I poke it". That, by any usual standards, is an avowal. We can suppose it is sincerely made. What are the reasons that support it? Well, they'd better include the desire that your GP has an accurate impression of your symptoms, your consequent intention not to mislead him by what you say, and *your awareness of when and where it hurts*. Maybe, I say again, there is some theoretically worthwhile sense in which your response to the doctor's question *expresses* your abdominal pain; maybe there isn't. But if there is, then whatever it is, it's not available to defuse the idea of privileged access. Rather, in order to understand your avowal as rationally performed, we need to presume that you have a kind of reliable awareness of the character of your pain that is denied to the doctor in advance of your expression of it in answer to his question.

The point is good for any performance, linguistic or otherwise, that is naturally conceived as giving *deliberate* expression to a state of mind. We can, without loss of generality, take any such expression to be the *first* occasion of giving vent to the state of mind in question. But then the relevant practical syllogism that rationalizes the subject's performance will have to include a belief about her state of mind as a precursor to its receiving outward expression. In short: to make practical rational sense of deliberate avowals, or any other form of deliberate expression of our mental states, involves advertent to beliefs of the subject about our own mental states which need to be regarded as in good standing if she is to be regarded as acting well, rationally speaking, but which cannot coherently be regarded as grounded in ways that are appreciable in principle by any observer of her performance.

## Knowing One's Own Meanings

Barry Smith's interesting and thoughtful chapter is given to the special case of self-knowledge of meaning. His chapter is concerned with what he calls the Reconciliation Problem, that is, the problem of reconciling self-knowledge of meanings, conceived as exhibiting the phenomena of privileged access, with the essential publicity of meaning—the idea not merely that anything one can mean by an expression is available to be meant by anyone else, but likewise that the fact of one's meaning it is publicly available too. In short, the problem is that of reconciling the publicity of what one means with the availability of one's meanings to oneself in the manner distinctive of intentional self-knowledge.

A crucial question, of course, is where the notion of the publicity of meaning springs from. It's an axiomatic common theme in Wittgenstein, Quine, Dummett, and Davidson. For Quine, and Davidson, the motivation is, *au fond*, metaphysical-naturalist:

there is, in their view, nothing for one's meaning something in particular by a particular expression to consist in if the expression's having that meaning does not make some distinctive, identifying impact on one's behavior. It is different for Wittgenstein, and following him Dummett: for these philosophers, publicity is a consequence of the idea that there is nothing to meaning save what is understood, and that understanding is a kind of knowledge-how, rather than knowledge-that—a complex of operational skills associated with the expression in question. Operational skills, in their very nature, have to be manifestable.

But these philosophers take quite different views of the other half of the reconciliation problem. Quine and, I believe, Davidson would simply reject any notion of privileged self-knowledge of meaning; all there is is the homophonic knowledge generated by disquotation into one's own idiolect. Wittgenstein was exercised by the phenomenology of the apparent transparency of one's own meanings—for instance, in the phenomenon of “grasping in a flash”.<sup>15</sup> But one senses that he regarded this as a puzzling, easily misunderstood phenomenon which gets in the way of a correct philosophical take on the notions of meaning and understanding, rather than a datum which any satisfactory account has somehow to integrate. Dummett, by contrast, regards what he calls the transparency of meaning to the thinker—more specifically, the transparency of Fregean senses—as an indispensable component of any satisfactory account of what it is to understand a language. On a view like Quine's, and Davidson's, I take it, there is no real reconciliation problem in the first place. The self-knowledge component of the problem is a chimera. It is on a Dummettian view that the problem is indeed acute, since on the one hand it is being asserted that meanings are by their very nature public, that understanding is essentially practical knowledge, and on the other that they are available to privileged access. There appears, on the face of it, no way that these claims can be made to cohere.

Smith's ambitious essay sets itself to provide a solution to this problem. His key move, however, is to resist the “exteriorization” of meaning that is the common theme running through Quine, Dummett, Wittgenstein, and Davidson. This leaves the status of the publicity of meaning somewhat in shadow. If meaning is indeed an inner psychological phenomenon, then the idea that it is also essentially public cannot be motivated by the kinds of considerations that moved those philosophers. Smith does not explain why he nevertheless accepts it, so that a reconciliation problem still arises. But he believes he can safeguard the publicity of meaning, even starting from a conception of meanings as interior.

Smith's leading idea is to apply to the purpose a notion of *epistemic entitlement* along the lines that I have myself tried to defend in various recent papers.<sup>16</sup> A key point for his project is one often stressed by John McDowell, that the phenomenology of understanding the speech of another involves *immediacy*—we does not, save

<sup>15</sup> See e.g. *Philosophical Investigations* §§138–99, 191 and 197.

<sup>16</sup> I pick up the discussion of this notion in Part IV of my Replies in this volume..

exceptionally, actively *interpret* the speech of another, but simply get her meaning by listening. Exactly what philosophical force is carried by this observation has always seemed somewhat moot to me. But Smith's proposal is that it is this: that in parsing the speech of another, one simply superimposes one's own idiolect onto her words, spontaneously understanding them as if produced by oneself. The crucial suggestion is then that if, as he takes the case to be, we are each of us defeasibly entitled to the general claim that anyone who uses an expression that is currency in one's idiolect means by it what one means oneself, then I can fetch up with knowledge of what you mean by a particular expression—to wit, exactly the same as I mean, which, our starting point was, I already know.

Smith's invocation of the notion of epistemic entitlement is somewhat briskly done, but I will not here delve into the question whether the case can be developed in such a way as to bear a clear structural analogy to other cases where, as I have argued, entitlements can be used to underwrite claims to knowledge. It is worth stressing, though, that in my own view, it is only *claims* to knowledge that can be underwritten in this way. The upshot of Smith's argument, if successful, ought to be, not that others' meanings are indeed available to us, but only that we are in position rationally to claim that they are. Whether that is a sufficiently robust sense of "publicity" to address any remaining reconciliation problem will depend on what Smith doesn't supply, namely a motivation for continuing to regard meanings as publicly available, once the "exteriorizing" move is rejected.

There may, however, seem to be a residual problem about the ability of Smith's kind of account to secure publicity in any worthwhile sense. If all goes well, each of us winds up entitled to claim to know that others share our respective idiolectic meanings. So your meanings, so I am entitled to claim, are available to me. But publicity had better involve that availability is reciprocal—that I should also be entitled to claim to know that *my* meanings are available to *you*. How does that work?

Here is a line that Smith might try. Suppose I am indeed entitled to claim to know that my meanings are the ones that you express when you use the same vocabulary. You too have the same entitlement. So you are entitled to claim to know that your meanings are shared by me. But these meanings, I can reflect, are mine—or so I am entitled to claim. So the meanings that you are entitled to claim to know are shared with others, including myself, will be—or so I should think—the very ones that I mean. So hey presto, I *am* entitled to claim to know that my meanings are available to you.

The main assumption of Smith's project is that we do indeed have privileged access to meaning. This has, of course, to relate to knowledge of *idiolectic* meaning—no one thinks they have privileged access to the meanings of Latin, or Dutch sentences. Smith, however, nowhere says what form he conceives self-knowledge of idiolectic meaning to take. Presumably it has to be propositional knowledge: the characteristic marks of immediacy, transparency, and authority engage opinions that so-and-so is the case. But then, what are the propositions in question like? It is at this point that one may begin to

have some doubt about the reality of the very phenomenon—of first-personal privileged access to idiolectic meanings—which Smith sets himself to reconcile with publicity. The knowledge in question has, presumably, to be metalinguistic—the propositions concerned have to be propositions *about* one’s idiolect, characterizing the meanings of its expressions, appropriately mentioned. But then the following point gives concern: there is no essential role for such knowledge in rationalizing one’s uses of the idiolect. Rationalizing the use of token expressions in the idiolect will require citation of various germane beliefs, desires, and intentions, and the citation of states of understanding. But understanding is not *metalinguistic* knowledge: to understand a language is not to have knowledge *about* it but to be able to use it to say what, modulo one’s other reasons, it is rational for one to want to say. The process of deliberation that leads up to the decision to say something need not, on pain of incompleteness, involve any semantic beliefs—any more than it needs syntactic ones. Of course the point remains that a normal adult speaker will typically have many beliefs of a semantic character. But they will be beliefs about the proper use of English, or French. To make rational sense of my use of a particular French word on a particular occasion, it may well be necessary to ascribe to me a metalinguistic belief about its meaning; and the reasons leading to my choice of the word may well involve such a belief. But the metalanguage in which these beliefs are formulated, if it is to be the language in which I articulate my reasons, had better be my idiolect. What is quite obscure is what role there is for metalinguistic beliefs *in* my idiolect *about* my idiolect. But such, it seems, would have to be the character of the beliefs that Smith regards as setting the problem of reconciliation.

The first-personal knowledge that Smith’s chapter is premised upon thus begins to seem fugitive on close inspection. Knowledge of one’s intentional states in general, of course, does involve knowledge of their content. But this is knowledge of *mental* content, not linguistic content. My realization that I was just now mistakenly thinking that tomorrow is Sunday does not come to me idiolectically garbed, as it were, and awaiting identification of its content via an application of my knowledge of what thought is expressed by “Tomorrow is Sunday” in my idiolect. Knowledge of the contents of one’s thoughts does not proceed through knowledge *about* one’s idiolect. But if the latter is needed to explain neither how one expresses oneself, idiolectically, nor how one knows what one thinks, what is it for? And why suppose we have it?

If this line of thought is correct, then we must revisit the somewhat negative assessment, reached at the end of Part I of these replies, of Paul Horwich’s idea of meanings as constituted in regularities of use, formulable in terms of ideal laws of which ordinary speakers may be quite unaware. It does seem plausible that the competent, purposive use of language must draw on some form of knowledge of meanings that is available to a subject at an intentional level—that there is a sense in which the subject “chooses her words”, in the light of her beliefs and goals, and no corresponding sense in which, for example, she “chooses a syntax” to frame the way she expresses herself. The syntax *shapes* her linguistic competence, but is not drawn upon, as personal-level

information, in its routine exercise. I canvassed it as an objection to Horwich that meaning seems to be different in just this respect: that if we are to explain a linguistic act as rational, we will need to cite the subject's understanding of it, alongside her other beliefs and desires, in a suitable practical syllogism, and that Horwich's account seems therefore to err in placing meanings beyond the sphere of ordinary personal knowledge. But the upshot of the foregoing consideration of Smith's project is that we have no plausible model of what knowledge of meaning, as personal self-knowledge, might consist in. The model that thrusts itself forward is essentially metalinguistic, and it cannot be knowledge of that character that the rational use of language essentially draws upon.

So is Smith's project misconceived, and is the objection to Horwich disarmed? Maybe. The question, I think, is whether some other, personal-level model of knowledge of meaning can be given to do justice to the idea that knowledge of one's own meanings somehow enters into the rationalization of intentional linguistic activity, at a personal level but in, as it were, quotation-free form, without taking the shape of propositional metalinguistic knowledge. If any such idea can be made coherent and plausible, there will be a question whether such knowledge comes within the scope of ordinary psychological self-knowledge of the kinds relevant here. If it does, the objection to Horwich's approach will stand, and Smith's issue about reconciliation with publicity will arise.

## Bibliography

- Bar-On, D. 2004 *Speaking My Mind*, Oxford, Clarendon Press.
- Bilgrami, A. 2006 *Self-Knowledge and Resentment*, Cambridge, Mass., Harvard University Press.
- Boghossian, P. A. 1989 "The Rule-Following Considerations", *Mind* 98/392, pp. 507–49.
- McDowell, J. 1998 "Response to Crispin Wright", in C. Wright, B. C. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds*, Oxford, Oxford University Press, pp. 47–62.
- Wright, C. 1992 *Truth and Objectivity*, Cambridge, Mass., Harvard University Press.
- Wright, C. 1998 "Self-Knowledge: The Wittgensteinian Legacy", in C. Wright, B. C. Smith, and C. Macdonald (eds.) *Knowing Our Own Minds*, Oxford, Oxford University Press, pp. 15–45.
- Wright, C. 2002 "What could Anti-Realism about Ordinary Psychology Possibly Be?", *The Philosophical Review* 111, pp. 205–33.