



## Knowing Our Own Minds

Crispin Wright (ed.) et al.

<https://doi.org/10.1093/0199241406.001.0001>

Published: 19 October 2000

Online ISBN: 9780191598692

Print ISBN: 9780199241408

Search in this book

CHAPTER

# 1 Self-Knowledge: The Wittgensteinian Legacy

Crispin Wright

<https://doi.org/10.1093/0199241406.003.0002> Pages 13–46

Published: October 2000

## Abstract

The central philosophical problem of self-knowledge is to account for, why, in an important basic class of cases, a subject's impressions of her own mental states are both groundless and authoritative, and for why those mental states are transparent to her. The Cartesian conception of the mind is best diagnosed as one such account, on the basis of the idea that one has a kind of observational access to one's inner states. The chapter reviews Wittgenstein's opposition to the Cartesian conception in philosophical investigations and raises the question: With what did he think it should be replaced? It is argued that the exegetically correct answer is: With nothing. Rather, the very question is a paradigm, in Wittgenstein's view, of what he regards as philosophy's misguided tendency to quest for explanation. The chapter concludes with a resumé of some of the discomforts, which this quietist position provokes. It also includes discussions of the possibilities for a successful expressivist treatment of psychological avowals, of immunity to error through misidentifications, and responds to John McDowell's criticisms of some of the author's earlier writings on self-knowledge and on Wittgenstein's views.

**Keywords:** [authority](#), [Cartesianism](#), [expressivism](#), [groundlessness](#), [immunity to error through misidentification](#), [transparency](#), [Wittgenstein](#)

**Subject:** [Epistemology](#), [Philosophy of Language](#), [Philosophy of Mind](#)

**Collection:** [Oxford Scholarship Online](#)

It is only in fairly recent philosophy that psychological self-knowledge has come to be seen as problematical; once upon a time the hardest philosophical difficulties all seemed to attend our knowledge of others. But as philosophers have canvassed various models of the mental that would make knowledge of other minds less intractable, so it has become unobvious how to accommodate what once seemed evident and straightforward—the wide and seemingly immediate cognitive dominion of minds over themselves.

My programme in this chapter involves characterizing this dominion with some care. We need to be as clear as possible why one form of traditional thinking on the matter has seemed so attractive—even unavoidable—and what a satisfactory account of the issues in this region has to accomplish. However, my underlying and primary

concern is with the later Wittgenstein's contribution to the question. Ultimately I think we are provided with a most vivid illustration—and can perhaps gain an insight into the intended force—of something which I do not think has so far been very well understood: the *anti-explanatory* motif that runs through the pronouncements on philosophical method occurring in the *Philosophical Investigations*.

I

People can be variously deluded about themselves: self-deceived about their motives, for instance, or overly sanguine, or pessimistic, about their strengths of character and frailties. But it is none the less a truism that for the most part we know ourselves best—better than we know others and better than they know us.

p. 14 In one kind of case, the explanation of this would seem straightforward. It is (merely) that our own presence is, for each of us, a constant ↪ factor in the kind of situation, usually but not always social, in which the evidence emerges which bears on various of our psychological characteristics. No one else is so constantly around us. So no one else observes as much of us or is as much observed by us. Selves have the best evidence about themselves.

Evidently, however, this form of explanation of the truth in the truism can run only in cases where one's own and another's knowledge of oneself must draw on the same kind of evidence. So it is restricted, it would seem, to broadly dispositional characteristics like honesty, patience, courage, and conceit—cases where there is no essential self/other asymmetry in the means of knowledge. And this is not, of course, the most salient type of case. In the most salient type of case, we do not merely know ourselves best, but also *differently* from the way in which we know others and they know us. The distinction is complicated, admittedly, by the fact that many apparently dispositional psychological characteristics are distinctively manifested not by raw behaviour, as it were, but by psychological performance in respects that may themselves exhibit self/ other epistemological asymmetries. Conceit, for instance, will be, *inter alia*, a disposition to form certain kinds of belief. It remains that the type of case that sets our problem is that which gives rise to the phenomenon of *avowal*—the phenomenon of authoritative, non-inferential self-ascription. The basic philosophical problem of self-knowledge is to explain this phenomenon—to locate, characterize, and account for the advantage which selves seemingly possess in the making of such claims about themselves.

The project will be conditioned by whatever more precise characterization we offer of the target phenomenon. It seems safe to suppose that we must begin by distinguishing two broad classes of avowal. The first group—what I will call 'phenomenal avowals'—comprise examples like 'I have a headache', 'My feet are sore', 'I'm tired', 'I feel elated', 'My vision is blurred', 'My ears are ringing', 'I feel sick', and so on. Such examples exhibit each of the following three marks.

First, they are *groundless*. The demand that somebody produce reasons or corroborating evidence for such a claim about themselves—'How can you tell?'—is always inappropriate. There is nothing they might reasonably be expected to be able to say. In that sense, there is nothing upon which such claims are based.

p. 15 Second, they are *strongly authoritative*. If somebody understands such a claim, and is disposed sincerely to make it about themselves, that is a guarantee of the truth of what they say. A doubt about such a claim has to be a doubt about the sincerity or the understanding of the one making it. Since we standardly credit any interlocutor, in the absence ↪ of evidence to the contrary, with sincerity and understanding, it follows that a subject's actually making such a claim about themselves is a criterion for the correctness of the corresponding third-personal claim made by someone else: my avowal that I'm in pain must be accepted by others, on penalty of incompetence, as a ground for the belief that I am.

Finally, phenomenal avowals exhibit a kind of *transparency*. Where P is an avowal of the type concerned, there is typically something absurd about a profession of the form, ‘I don’t know whether P’—don’t know whether I have a headache, for instance, or whether my feet are sore. Not always: there are contexts in which I might be uncertain of a precondition—for instance, whether I have feet. But in the normal run of cases, the subject’s ignorance of the truth or falsity of an avowal of this kind is not, it seems, an option.

None of the examples listed is an avowal of a *content-bearing* state. It is the hallmark of the second main group of avowals—what I shall call ‘attitudinal avowals’—that the psychological characteristics, processes, and states which they concern are partially individuated by the propositional content, or intentional direction, which informs them—for instance, ‘I believe that term ends on the 27th’, ‘I hope that noise stops soon’, ‘I think that professional philosophers are some of the most fortunate people on earth’, ‘I am frightened of that dog’, ‘I am thinking of my mother’. In order to see what is distinctive about an author’s relation to avowals of this kind, we need first to take account of the fact that such claims can also be made as part of a process of *self-interpretation*—in the kind of context where we say that we have *learned* about our attitudes by finding that certain events cause us pleasure, for instance, or discomfort. Consider the following passage from Jane Austen’s *Emma*:

Emma’s eyes were instantly withdrawn; and she sat silently meditating in a fixed attitude, for a few minutes. A few minutes were sufficient for making her acquainted with her own heart. A mind like hers, once opening to suspicion, made rapid progress. She touched—she admitted—she acknowledged the whole truth. Why was it so much the worse that Harriet should be in love with Mr. Knightley than with Mr. Churchill? Why was the evil so dreadfully increased by Harriet’s having some hope of return? It darted through her, with the speed of an arrow, that Mr. Knightley must marry no one but herself.<sup>1</sup>

p. 16 Here Emma has just been told of the love of her protégée, Harriet, for her—Emma’s—bachelor brother-in-law, a decade older than Emma, a frequent guest of her father’s, and hitherto a stable, somewhat avuncular part of the background to her life. She has entertained no thought of him as a possible husband. But now she realizes that she strongly desires that he marry no one but her, and she arrives at this discovery by way of surprise at the strength and colour of her reaction to Harriet’s declaration, and by way of a few minutes’ reflection on that reaction. She is, precisely, not moved to the realization immediately; it dawns on her as something she first suspects and *then* recognizes as true. It *explains* her reaction to Harriet.

In such self-interpretative cases, none of the three features we noted of phenomenal avowals is present. There is no Groundlessness: the subject’s view is one for which it is perfectly in order to request an account of the justifying grounds. There is no Strong Authority: mere sincerity and understanding will be no guarantee whatever of truth—it is for Jane Austen to stipulate, as it were, that Emma’s self-discovery is the genuine article, but in any real context such a conclusion could be seriously mistaken. Finally, there is no Transparency: within a context of self-interpretation, it is in no way incongruous if the subject professes ignorance of particular aspects of her intentional psychology. But what it is vital to note for our present purpose is that such self-interpretative cases, although common, cannot be the *basic* case. For the body of data on which self-interpretation may draw is not restricted to recollected behaviour and items falling within the subject-matter of phenomenal avowals. When Emma interprets her reaction to Harriet’s declaration as evidence that she herself loves Knightley, there is an avowable ground—something like ‘I am disconcerted by her love for that man and, more so, by the thought that it might be returned’—which is a *datum for*, rather than a *product of*, self-interpretation. Self-interpretation, that is to say, will typically draw on non-inferential knowledge of a basic range of attitudes and intentionally characterized responses. These will not be distinguished, I think, from non-basic, interpretative cases by any generic features of their content; rather, they will reflect matters which, for the particular subject in the particular context, happen to require no interpretation to be known about—matters which are precisely *avowable*. It is these basic examples which comprise the attitudinal avowals.

p. 17

Such avowals will have the same immediacy as phenomenal avowals, and will exhibit both groundlessness and transparency—groundlessness rather trivially, in so far as, any interpretational basis having been excluded, there will naturally be nothing a subject can say to justify such a self-ascription; transparency in the sense that, except where the matter is one of interpretation, we think a subject ought to know without further ado what she believes, or desires, etc., so that any profession ↵ of ignorance or uncertainty, unless coupled with a readiness to allow that the matter is not basic but calls for (self-) interpretation, will seem perplexing. However, attitudinal avowals do not exhibit the strong authority of phenomenal avowals: to the extent that there is space for relevant forms of self-deception or confusion, sincerity-cum-understanding is no longer a guarantee of the truth of even basic self-ascriptions of intentional states. Any avowal may be discounted if accepting it would get in the way of making best sense of the subject's behaviour. But with attitudinal avowals, it is admissible to look for explanations of a subject's willingness to assert a bogus avowal other than those provided by misunderstanding, insincerity, or misinterpretation. This is the space occupied by the ordinary notion of self-deception; but the more general idea is just that we can be caused to hold mistaken higher-order beliefs in ways—wishful thinking, for instance—which do not go through misguided self-interpretative inference.

p. 18

It is striking that attitudinal avowals would appear to exhibit a form of weak authority nevertheless: that is, they provide criterial—empirically assumptionless—justification for the corresponding third-person claims. Since it cannot be attributed, as with phenomenal avowals, to the fact of sincerity-cum-understanding guaranteeing truth, what does this weak authority consist in? It might be suggested that it is nothing other than the presumptive acceptability of testimony generally. And certainly that proposal would be enough to set our problem: for the presumptive acceptability of *original* testimony—testimony for which the source is not itself testimony—extends no further than to subject-matters which an informant is deemed competent to know about. So the question would recur: how is it possible for subjects to know about their intentional states in ways that involve no consideration of the evidence on which a third party must rely? Actually, however, I think the suggestion is wrong. What distinguishes the presumptive acceptability of attitudinal avowals from anything characteristic of testimony generally is that the authority which attaches to them is, in a certain sense, *inalienable*. There is no such thing as showing oneself chronically unreliable in relation to the distinctive subject-matter of attitudinal avowals. I may have such poor colour vision that you rightly come to distrust my testimony on matters of colour. I may, unwittingly, have a very bad memory, and, learning of this, you may rightly come to a state of wholesale suspicion about my testimony on matters of personal recall. But no corresponding wholesale suspicion concerning my attitudinal avowals is possible. You may not suppose me sincere and comprehending, yet chronically unreliable, about what I hope, believe, fear, and intend. Wholesale suspicion about my attitudinal avowals—where it is not a ↵ doubt about sincerity or understanding—jars with conceiving of me as an intentional subject at all.

## II

Both groups of avowals exhibit a further feature which it is worth attending to briefly.<sup>2</sup> In a famous passage in the *Blue Book*, Wittgenstein writes as follows

There are two different cases in the use of the word 'I' (or 'my') which I might call the 'use as object' and 'the use as subject'. Examples of the first kind of use are these: 'my arm is broken', 'I have grown six inches', 'I have a bump on my forehead', 'the wind blows my hair about'. Examples of the second kind are: 'I see so-and-so', 'I hear so-and-so', 'I try to lift my arm', 'I think it will rain', 'I have toothache'. One can point to the difference between these two categories by saying: the cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error or as I should rather put it: The possibility of an error has been provided for . . . [but] it is as impossible that in making the statement 'I have toothache', I should have mistaken another person for myself, as it is to moan with pain by mistake, having mistaken someone else for me.<sup>3</sup>

The characteristic to which Wittgenstein here calls attention is often called ‘immunity to error through misidentification’.<sup>4</sup> In a large class of cases, when someone makes a subject–predicate claim, they may mistake or misidentify the subject in a way for which, it appears, there is no provision in the case of an avowal. If I see someone running along the beach, and, taking it to be my colleague NN, say, ‘NN will catch up with us in a minute’, I may be mistaken in ways which correspond either to the subject identified or to the predication I make: it may be that the character approaching us is not NN at all, and it may be that he will turn away before he reaches us. But if I avow my indifference to a forthcoming ballot, for instance, then there is a provision for correction only for my predication—I can’t be mistaken about whom I’m making that predication of.

p. 19 Shoemaker writes that the absence of the possibility of misidentification ‘is one of the main sources of the mistaken opinion that one cannot be an object to oneself, which in turn is a source of the view  $\hookrightarrow$  that “I” does not refer’.<sup>5</sup> The line of thought he has in mind runs from the impossibility of *misidentifying* myself to the conclusion that, in reflection about myself, there is no identification of *any* object, that such reflection is not object-directed at all, and that ‘I’ accordingly has no referential function.

Wittgenstein may, indeed, have been inclined to such an idea—the passage which I quoted continues: ‘To say, “I have pain” is no more a statement *about* a particular person than moaning is.’<sup>6</sup> But whether he was so inclined or not, it is simply a mistake to suppose that where there is no fallible identification, there is no reference either. It will be clear what’s wrong with this idea once we have a better characterization of what immunity to error through misidentification (IEM) consists in. And we shall also thereby learn why, in the present context, the phenomenon need not be of primary concern.

The key point is that IEM is not a characteristic which a statement has simply in virtue of its subject-matter. It depends upon the kind of ground which a speaker has for it. Specifically, the ground has to be such that in the event that the statement in question is somehow defeated, it cannot survive as a ground for the corresponding existential generalization. Consider again the case of the figure on the beach, and suppose that, having asserted ‘NN will catch up with us soon’, I turn to see NN standing just a little in front of me. In that case, my original thought is defeated; but the basis for it survives as a ground for the claim, ‘Someone will catch up with us soon’. Or suppose I catch sight of my reflection passing a shop window, and I say to myself, ‘My hair is blowing in the wind’. If it becomes apparent that the reflection I saw was not mine but someone else’s, the basis for my claim will still remain as a ground for the existential generalization, ‘Someone’s hair was blowing in the wind’. A claim, made on a certain kind of ground, involves immunity to error through misidentification just when its defeat is *not* consistent with retention of grounds for existential generalization in this kind of way.

p. 20 Now, avowals do characteristically so behave. If an avowal of mine is somehow defeated, there is no question of my original entitlement surviving, without my gathering additional information, in such a way that I may justifiably claim that someone—though not me—exhibits the property which I avowed. However, two points are evident in the light of the characterization just offered. First, the idea that Wittgenstein’s ‘as subject’ uses of ‘I’ are somehow shown to be non-referential by their  $\hookrightarrow$  having IEM should have been strangled at birth by the reflection that a similar immunity is characteristic of many *demonstrative* claims, in which case there is of course no question but that reference to an object is involved. If I see an object hurtling towards us and say, ‘That thing is approaching very fast’, there is no way in which that claim can be defeated yet my original grounds for it survive as grounds for the claim, ‘Something is approaching very fast’. Second, it is clear that many *non-psychological* claims about the self can exhibit the same feature. Consider, for instance, my assertion ‘My hair is blowing in the wind’ when based not on a shop-window reflection but on certain characteristic feelings of the scalp and face and my auditory sensations. Or consider ‘The bedclothes have fallen off my leg’, uttered by an unwitting recent amputee. In neither case is there the possibility of a justified fall-back to an appropriate existential generalization if the claim is defeated. Immunity to error through misidentification is not a distinctive characteristic of *psychological* self-ascriptions or, more specifically, of avowals.

p. 21

When I said a moment ago that a better characterization of IEM would teach us why this feature need not be of special concern to us in the present context, I was not merely anticipating this point. That the phenomenon extends to examples like those just given will seem to be quite consistent with its still betokening something essential about avowals once the plausible thought occurs that the *source* of IEM in non-psychological first-personal claims is always their being based on avowable psychological matters (for instance, my sensations of scalp and face and the rushing in my ears, or the amputee's sensations as of a draught around his foot). The IEM of non-psychological self-ascriptions, when they have it, is presumably to be viewed as an inheritance from their basis in an underlying possible avowal. It is not the fact that IEM is not the exclusive property of psychological claims which entitles us to bracket the phenomenon in the present context, but rather the reflection that avowals' exhibition of it is a *derived* feature, as it were —effectively a consequence of respects in which we have already noted their distinction from third-personal psychological claims. Specifically it is a consequence of their being groundless while the corresponding third-personal claims demand evidential support. For if an avowal, 'I am  $\phi$ ', did not exhibit IEM, then its defeat would be consistent with the subject's retention of an entitlement to the corresponding existential generalization—'Someone (else) is  $\phi$ '—which could then be asserted *groundlessly*. But to suppose that such a claim could be both admissible and groundless would clash with the original asymmetry. Claims about the psychological states of others are acceptable only when grounded,  $\hookrightarrow$  one way or another; that goes both for particular such claims and for generalizations of them as well.

### III

It is natural to wonder what, if any, independent, general characterization may be possible of the psychological distinction marked by the contrast between phenomenal—and attitudinal—avowals, as outlined. Here I can only observe that neither of two initially suggestive proposals seems to be quite right. Familiarly, the *Investigations* repeatedly counsels against construing understanding, hoping, fearing, intending, etc. as mental states or processes. Wittgenstein's idea was not, of course, that there are no such things, strictly, as mental processes, or the states that would constitute their end-points, so to speak, but only that understanding, etc. will be misunderstood if assimilated to them. So the suggestion invites consideration that the distinction he is making, between mental events and processes strictly so termed—turns in consciousness, if you like—and other psychological states, corresponds nicely to that between phenomenal and attitudinal avowals: that phenomenal avowals register states or processes of mind which Wittgenstein would be content to describe as such, whereas attitudinal avowals mark cases where danger lurks in that description.<sup>7</sup> The exegetical question is worth marking, but I shall not pursue it here.<sup>8</sup> Suffice it to say that some things which are surely 'turns in consciousness' in anyone's book—for instance, a lover's face passing before your mind's eye, or the thought occurring to you that you should have phoned home twenty minutes ago, or being startled by an opponent's outburst at the umpire—such cases are, unlike any of our prototypes of the phenomenal-avowable, also *contentful*; moreover, the proper description of such an episode is, at least in some cases, something about which a subject might just conceivably be self-deceived.

p. 22

Another mark of many phenomenal-avowable states—and certainly of each example I cited originally—is that, at least as we ordinarily suppose,  $\hookrightarrow$  creatures may be subject to them while *having no concept of them*: a dog can be tired, or afraid, or have an itch, without having any concept of those states. But this distinction too seems not to coincide quite cleanly with that between the phenomenal- and the attitudinal-avowable. Some phenomenal-avowable conditions do appear to require the subject to have a concept of them. I doubt, for instance, that a dog can have *Auld Lang Syne* run through its head, although that is presumably a process amenable to strongly authoritative avowal by a suitably endowed subject—though I suppose it depends on what you regard as required by an understanding of 'I am imagining a rendition of *Auld Lang Syne*'. Conversely, we tend quite freely to ascribe attitudinal states to (sufficiently intelligent, adaptable) creatures—primates and dogs, for instance—who presumably have no concept of such states.

In any case, we have sufficient of a focus for our central question. The cardinal problem of self-knowledge is that of explaining *why* avowals display the marks they do—what it is about their subject-matter, and the subject's relationship to it, which explains and justifies our accrediting her sincere pronouncements about it with each of Groundlessness, Strong Authority, and Transparency in the case of phenomenal avowals, and with Groundlessness, Weak Authority, and Transparency in the case of attitudinal avowals. How is it possible for subjects to know these matters non-inferentially? How is it (often) impossible for them *not* to know such matters? And what is the source of the special authority carried by their verdicts?

## IV

---

There is a line of response to these questions that comes so naturally as to seem almost irresistible—indeed, it may even seem to ordinary thought to amount merely to a characterization of the essence of mind. According to it, the explanation of the special marks of avowals is that they are the product of the subject's exploitation of what is generally recognized to be a position of (something like) *observational privilege*. As an analogy, imagine somebody looking into a kaleidoscope and reporting on what he sees. No one else can look in, of course, at least while he is taking his turn. If we assume our hero to be perceptually competent, and appropriately attentive, his claims about the patterns of shape and colour within will exhibit analogues of each of the marks of phenomenal avowals:

p. 23

- (i) The demand that he produces reasons or corroborating evidence for his claims will be misplaced—the most he will be able to say  $\hookrightarrow$  is that he is the only one in a position to see, and that is how things strike him;
- (ii) Granted his proper perceptual functioning, it will be sufficient for the truth of his claims that he understands them and is sincere in making them; so for anyone who understands the situation, our hero's merely making such a claim will constitute a sufficient, though defeasible, reason for accepting its truth; and
- (iii) Where P is any claim about the patterns of shape and colour visible within, there will be no provision—bearing in mind our hero's assumed perceptual competence and attentiveness—for his intelligibly professing ignorance whether or not P.

The analogy isn't perfect by any means. In order to construct it, we have had to assume normal perceptual functioning and full attentiveness on the part of our observer. And no such assumption conditions our reception of other's avowals. But once into one's stride with this type of thinking, this difference will not seem bothersome. The line will be that in the *inner* observational realm, in contrast to the outer, there is simply no room for analogues of misperception or of oversight or occlusion—for the objects and features there are necessarily salient to the observing subject. Or, at least, they are so in the case where they are objects and features recordable by phenomenal avowals. In the case of the subject-matter of attitudinal avowals, by contrast, space for an analogue of misperception can and should be found—that will be what explains the failure of strong authority in those cases. In brief, this—Cartesian—response to the problem of avowals has it that the truth-values of such utterances are non-inferentially known to the utterer via her immediate awareness of events and states in a special theatre, the theatre of her consciousness, of which others can have at best only indirect inferential knowledge. In the case of phenomenal avowals, this immediate awareness is, in addition, infallible and all-seeing; in the case of basic attitudinal avowals, it is merely very, very reliable.

So presented, the Cartesian picture—of the transparency of one's own mind and, by inevitable contrast, the opacity of others'—emerges as the product of a self-conscious attempt at philosophical explanation. This may seem congenial to John McDowell's claim that 'We need to be seduced into philosophy before it can seem natural to suppose that another person's mind is hidden from us'.<sup>9</sup> McDowell recoils from the idea that

anything like the Cartesian picture might be part of ordinary unphilosophical thought. But I think he is wrong about this, the theoretical setting I have given to the picture notwithstanding. To be sure, it is unclear what should count as a 'seduction into philosophy'. But if every manifestation of the Cartesian picture is to rate as the product of such a seduction, then the seductive reach of philosophy is flatteringly wide. I do not imagine, of course, that people typically self-consciously follow through the train of thought I outlined. But we ought not to balk at the notion that no intellectual routine characteristically pursued by those in its grip should capture exactly the best reconstruction of why an idea appeals. The privacy of the inner world is a recurrent idea in literature.<sup>10</sup> It is arguably a presupposition of the whole idea of the continuation of one's consciousness after death. The thought of the undetectable inverted colour spectrum is something which can engage quite young children without too much difficulty. And in each of these cases, what comes naturally is essentially nothing other than the notion of a kind of privileged observation of one's own mind, which works, in the ways we have reviewed, to explain the first-third-person asymmetries in ordinary psychological discourse.

The privileged observation explanation is unquestionably a neat one. What it *does* need philosophy to teach is its utter hopelessness. One very important realization to that end is that nothing short of full-blown Cartesianism can explain the asymmetries in *anything like the same way*—there can be no scaled-down observational model of self-knowledge which preserves the advantages of the Cartesian account while avoiding its unaffordable costs. The problem is that the kind of authority I have over the avowable aspects of my mental life is not transferable to others: there is no contingency—or, none of which we have any remotely satisfactory concept<sup>11</sup>—whose suspension would put other ordinary people in a position to avow away on my behalf, as it were. So the conception of avowals as reports of inner observation is saddled with the idea that the observations in question are ones which *necessarily* only the subject can carry out. And once that conception is in place, others' means of access to the states of affairs which their subject (putatively) observes is bound to seem essentially second-rate by comparison, and to be open to just the kinds of sceptical harassment which generate the traditional problem of other minds—the unaffordable cost referred to.

## V

If this is right, then a deconstruction of the privileged observation solution to the problem of self-knowledge is the indispensable prerequisite for an overall satisfactory philosophy of mind. It seems to me that it was Wittgenstein who first accomplished such a deconstruction, and I shall here try briefly to defend a certain conception of the way the deconstruction goes. In essentials, what he does is to mount a two-pronged attack on the Cartesian picture, with the two prongs corresponding to the distinction between the two main kinds of avowals. It is the so-called private language argument—the batch of considerations that surface in §§243 to the early 300s in the *Investigations*—which targets the idea of phenomenal avowals as inner observation reports, while the corresponding conception of attitudinal avowals is challenged by the various phenomenological and other considerations which Wittgenstein marshals in the, as we may call them, 'not a mental process' passages recurrent throughout the text.<sup>12</sup>

We need some preliminaries about *Investigations* §258, the famous passage in which Wittgenstein suggests that the 'private linguist' lacks the resources to draw a distinction which is essential if the 'reports' that he logs in his diary are to have a truth-evaluable content: the distinction, namely, between what seems right to him and what is really right. It seems clear that such a distinction is called for if these 'reports' are to have anything of the objectivity implicit in the very idea of an observational report, the objectivity implicit in the idea of successful representation of some self-standing aspect of reality. But it is not obvious that such a distinction is in good standing only if it can be drawn *operationally*—only if, that is, the diarist has the resources for making principled, presumably retrospective, judgements about occasions when he has been ignorant or mistaken. It ought to be enough if what constitutes the fact of the correctness of a report and what constitutes the diarist's

impression of its correctness are not all the same thing, so to speak, even if no one can ever be in the position of ascertaining the one without the other.

p. 26 Some commentators have taken it that Wittgenstein has missed this: that his objection here is implicitly verificationist—implicitly demands *contrastive uses* for locutions corresponding to the two halves of the distinction, and thus in effect begs the question against the idea that a subject might be *infallibly* aware of aspects of her inner life which are nevertheless constitutively independent of that awareness in the manner which the *seems right/is right* distinction, properly understood, demands. This accusation seems to me mistaken. Any idea that the private diarist's jottings might be subject to a non-operational, but still valid, *seems right/is right* distinction comes into collision with an important aspect of Wittgenstein's discussion of the notion of intention. Broadly: the fact of an aspect of the diarist's inner world being as one of his reports states it to be demands, if it is to be constitutively independent of his impression of the fact, the existence of facts about what is required by the *semantics of the private language* which are likewise constitutively independent of his impressions of facts about them. Since the semantics of the private language are to be constituted in certain original intentions of the diarist, the upshot is that a similar constitutive independence is required between facts concerning what really complies with the diarist's original semantic intentions and his own subsequent impression of those facts, even under best conditions—conditions of the utmost lucidity, perfect recall, etc. But that is a sustainable demand only if the content of an intention is objectionably *platonized*, in a fashion which, after Wittgenstein's discussions (especially in *Investigations* §§600 ff.), we should know better than to do.

That, of course, is merely to sketch a line of interpretation which I have pursued before, but have no space to elaborate here.<sup>13</sup> But the sketch should be enough to set the stage for three objections which John McDowell has lodged,<sup>14</sup> to which I now turn.

## VI

The first objection is a perfectly natural one. Is there not a prior, rather obvious doubt about the whole idea that the semantics of a private language could be constituted in certain original intentions of its practitioner? We should ask, McDowell says,

p. 27 how an intention could be constitutive of a concept, as opposed to annexing an independently constituted concept to a word as what it is to express. The private linguist's semantical intention is supposed to be inwardly expressible by something like 'Let me call this kind of thing "S" in future'. But for this story even to seem to make sense, the classificatory concepts supposedly expressed, with the help of an inward focusing of the attention, by 'this kind of thing' would need to be at the linguist's disposal already; it cannot be something he equips himself with by such a performance. If a new classificatory concept can be set up by focusing on an instance, that is only thanks to the prior availability of a concept that makes the right focus possible, in the presence of the instance, by fixing what kind of classification is in question.<sup>15</sup>

McDowell's suggestion is that my putative difficulty with the idea of a report's complying with the semantics of the private language is a distraction: that the place at which the real difficulty is to be found is in the idea that the diarist could intentionally constitute a semantics for his language in the first place. For how exactly is the private ostensive definition supposed to work? If focusing the attention has to be mediated by concepts—surely one of the prime intended morals of Wittgenstein's discussion of ostensive definition—then a private ostensive definition could take place only in a context akin to the Augustinian setting introduced right at the start of the *Investigations*, wherein the trainee is assumed already to be a *thinker*—already to be master of a range of concepts of which, as a result of ostensive training, he will for the first time acquire means of expression. That won't do.

There is no doubt about the Wittgensteinian resonances of this line of thought. But McDowell doesn't see that there is nothing in it to disturb the essentials of the Cartesian view. The reason is that the self-directive role which Wittgenstein gives to his diarist in §258 is actually quite inessential to the putative upshot—the operation of a private language. We are not, after all—or had better not be—tempted to say that a similar difficulty must afflict normal, public ostensive teaching. We don't think that a child must somehow already be equipped with concepts of colour if he is to benefit from a normal ostensive training in the use of colour words—that, indeed, is the point of the *Investigations* contrast between ostensive definition proper and what Wittgenstein calls ostensive *teaching*.<sup>16</sup> But if that is right, there is in general no barrier to the idea that something conceptless—a colour-sensitive machine, say—might serve to do the teaching in place of a normally colour-concepted adult. Better, there ought to be no difficulty in the idea that a blind man, taking himself to have some form of prosthetic indication of variations in colour, might successfully introduce a normally sighted child to a range of colour concepts which he himself lacked. But *that is* essentially the model that Cartesianism, in so far as it figures in ordinary thought, offers of the teaching of sensation concepts. According to that model, I am screened from the inner goings-on when my child exhibits pain behaviour. But I take it that the behaviour gives me a kind of prosthesis—that it betokens inner phenomenal saliences which I can accordingly train him to vocalize, and thereby to conceptualize.

p. 28 In short, the diarist/private linguist doesn't have to be an autodidact. So no essential difficulty with the Cartesian conception of sensation language is disclosed by drawing out problems inherent in the idea of such autodidacticism. At some point, the Cartesian should say, a subject's competence in the linguistic routines in which, in both the inner and the outer cases, he is trained, will amount to the possession of concepts. And the identity of the concepts then possessed will supervene on the linguistic intentions of the subject: on the patterns of use which he will be willing to uphold. So there is no alternative but that a discussion of the idea of a subject's unwitting and undetectable departure from a prior intention must take centre stage if Wittgenstein's deconstruction is to succeed.

## VII

McDowell's second objection is that the interpretation I offer winds up saddling Wittgenstein with the denial of certain platitudes. My claim was that we should elicit from Wittgenstein's discussion of intention and cognate concepts the moral that, as I put it in the discussion to which McDowell is responding, 'there is nothing for an intention, conceived as determining subsequent conformity and non-conformity to it autonomously and independently of its author's judgements on the matter, to be'.<sup>17</sup> McDowell glosses this as 'there is nothing but platonistic mythology in the idea that an intention determines what counts as conformity to it independently of its author's judgements'. He continues:

So [Wright's] thesis is that what the private linguist needs in his semantical intentions is something that cannot be true of any intentions at all, on pain of platonism.

But suppose I form the intention to type a period. If that's my intention, it is settled that only my typing a period will count as executing it. Of course I am capable of forming that intention only because I am party to the practices that are constitutive of the relevant concepts. But if that is indeed the intention which—thus empowered—I form, nothing more than the intention itself is needed to determine what counts as conformity to it. Certainly it needs no help from my subsequent judgements. (Suppose I forget what a period is.) So there is something for my intention to type a period, conceived as determining what counts as conformity to it autonomously and independently of my judgements on the matter, to be: namely, precisely, my intention to type a period. An intention to type a period is exactly something that must be conceived in that way. This is commonsense, not platonism.<sup>18</sup>

p. 29 This response is indeed commonsensical, but it completely misses the point. Of course, if I form the intention that P, what will comply with it is only and exactly the bringing it about that P; and it will typically be a matter independent of my subsequent judgements whether or not just that has been effected. The role of subsequent judgement is indeed not to mediate somehow in the connection between the content of an intention and its execution—granted, that idea jars with the very idea of an intention having a determinate content—but rather to enter into the determination of *what* the content of an anterior intention is to be understood as having been. If I form the intention to type a period, then, sure, only typing a period will do. The anti-platonist point is rather that there is nothing for my intention's having had *just that content* to consist in, if the fact has to be constitutively independent of anything which I may subsequently have to say about compliance or non-compliance with the intention, or about what its content was.

The platonist mythology is a mythology of such constitutive independence. And it is exactly what the private linguist needs if the required seems right/is right distinction is to exist in the kind of way we are currently considering. Against this platonism I want to set what I take to be an idea of Wittgensteinian authorship, although it is of course very familiar from the writings of Davidson: that the content of a subject's intentional states is not something which may merely be *accessed*, as it were indirectly, by interpretative methods—rather as, on a platonistic philosophy of mathematics, a good proof is merely a means of *access* to a mathematical truth—but is something which is intrinsically sensitive to the deliverances of best interpretative methodology. That is a methodology which in principle must include within its conspectus the whole sweep of a subject's sayings and doings, including future ones, without bound.

## VIII

---

The foregoing concerns the first prong of Wittgenstein's deconstruction of the Cartesian view—the anti-private language argument. The second prong, according to my reading, concerns certain special difficulties in the idea that attitudinal avowals describe introspectable mental occurrences. The suggestion of mine to which McDowell principally takes exception here is that one such special difficulty concerns the answerability of ascriptions of intentional states, like expectation, hope, and belief, to aspects of a subject's outward performance that may simply *not be available* at the time of avowal. If an expectation, say, were a ↴  
p. 30 determinate, dated occurrence before the mind's eye, then in any particular case it would either have taken place or not, irrespective of how I subsequently went on to behave. So we ought to be guilty of a kind of conceptual solecism if we hold claims about expectation to be answerable to subsequent sayings and doings in a fashion broadly akin to the way in which the ascription of dispositional states is so answerable. Yet that is exactly what we actually do.

The point was meant to be, then, that the conception of attitude avowals as reports of inner observation enforces a view of their subjectmatter which is at odds with another, fundamental feature of their grammar—their 'disposition-like theoreticity'.<sup>19</sup>

McDowell—this is his third objection—thinks this is wrong, both in substance and as a reading of Wittgenstein. There cannot, he contends, be a difficulty of this kind if there are examples of unquestioned phenomena of consciousness which nevertheless bear the same kind of internal connections to the outer. Yet there surely are. The coming of a picture before the mind is an episode in consciousness if anything is; yet the relation between the picture and the real scene it pictures, if there is one, is presumably comparable in relevant respects to that between an expectation and its fulfilment. So intentional states can indeed be mental occurrences—indeed, we noted this earlier. And equally, non-intentional occurrences—including the proto-example of pain itself—are in a similar case: if a pain isn't an inner scenario, then nothing is—yet the ascription of pain is, likewise, answerable in complex but constitutive ways to the subject's outward performance.

McDowell lodges this objection in the context of a reading which has me attempting a rigorous and exhaustive division between strict phenomena of consciousness—Rylean ‘twinges and stabs’—which, McDowell thinks I think, involve no intentionality, and whose occurrence is independent of the conceptual resources of the subject, and intentional states and processes which are never, properly speaking, occurrent phenomena of consciousness. That is a misreading, and the second prong does not require any such clean distinction. All that is required is that there be a difficulty along the indicated lines for at least one wide class of avowals. Still, the objection remains: isn’t any such difficulty defused by the perfectly valid observation that items which certainly ought to be counted as mental *occurrences* do, after all, sustain internal relations to the outer?

p. 31 I don’t think so. For one thing, remember that the dialectical setting is one in which even the claim of twinges and stabs to count as pure occurrences before an inner eye is under question. Conceived as by the Cartesian, a pain and the behaviour which expresses it are *quite distinct existences*, the one visible only to the subject, the other in public view. There is an incongruity, on that conception, in the conceptual linkage between the two which conditions ordinary psychological discourse. It is an instance of a more general incongruity which the sceptic quite rightly—when mentality is conceived as on the Cartesian view—finds in our empirically ungrounded reliance upon what is outer as a guide to subjects’ mental lives.

The real counter to McDowell’s objection, however, comes with a second consideration. When an image, or picture, comes before my mind, it presumably cannot constitute a more explicit or substantial presence than the coming of a real physical picture before my physical eye. And when the latter happens, it is of course consistent with my being in full command of all manifest features of the object that I remain ignorant precisely of its intentionality—of what it is a picture of. I want to say that, analogously, in the sense in which an image or mental picture can come before the mind, its intentionality cannot. Wittgenstein himself is making this point when he asks, ‘What makes my image of him into an image of *him*?’<sup>20</sup> and answers, having said that ‘the same question applies to the expression “I see him now vividly before me” as to the image’, ‘Nothing in it or simultaneous with it’. But this aspect—the intentional content—of expectation, belief, and their kin still falls firmly within the province of the non-inferential authority which we accord to attitudinal avowals. And it remains that the model of inner observation is bankrupt to explain the fact.

McDowell writes:

Wright’s reading actually puts it in doubt whether anything could be an occurrent phenomenon of consciousness. . . . The kinds of connection that raise the problem for intentional states are, as I said, connections involving the ‘normative’ notion of accord. But such connections are a species of a wider genus, that of internal relations. If Wright’s a priori argument worked, it would have to be because nothing introspectable could sustain internal relations to anything outer.<sup>21</sup>

p. 32 I have already said that I think there is a doubt whether occurrences in a Cartesian theatre of consciousness could sustain the kind of conceptual connections with outward performance which sensations, as we actually conceive of them, do sustain. But this ought to be consistent with the ordinary idea that a pain, say, is an item for introspection. It is a misunderstanding of the argument if it seems that, were it successful, nothing introspectable could sustain internal relations to anything outer. It all depends on the character and source of the internal relations in question. Both a sunburned arm, if I may borrow Davidson’s example, and a triangle can be presented as ordinary objects of observation, and each sustains, *qua* presented under those particular respective concepts, certain internal relations: the sunburned arm to the causes of its being in that condition, and the triangle to, for instance, other particular triangles. And the point is simply that while identification of the triangle as such can proceed in innocence of its internal relations of the latter kind—maybe the subject has no knowledge of the other triangles at all—recognition of the sunburned arm as just that cannot proceed in like innocence, but demands knowledge that its actual causation is as is appropriate to that mode of presentation of it. The general form of the point that I take Wittgenstein to be making in the second prong of his attack is that the internal relations to the outer which are constituted in the intentionality of psychological

items, of whatever sort, are all of the latter, sunburn-style kind. Hence there is indeed a standing puzzle in the idea that an appropriate characterization of them, incorporating such intentionality, is somehow vouchsafed to their subject by something akin to pure observation, *a fortiori* in the idea that it is the privileged character of this observation which underlies first-personal authority about such states. This thought, it seems to me, continues to be impressive.

p. 33 Interpreting Wittgenstein on these issues is a subtle and difficult matter, and not just because of the subtlety and difficulty of his thinking. McDowell's own positive reading<sup>22</sup> involves construal of the target of the anti-private language argument as a version of the Myth of the Given, transposed to the inner sphere; and he holds that the correct interpretation of the 'not a mental process' passages should still be consistent with the capacity of items such as meanings, intentions, and their kin to 'come before the mind', as he likes to say.<sup>23</sup> I must defer further consideration of his proposals. Here, I merely want to say at least a minimum to explain my continuing conviction that the reading of  $\hookrightarrow$  Wittgenstein I have defended should lead one to recognize deep incoherences in the Cartesian response to our problem—incoherences that are prior to its inordinate sceptical costs. And if what I said earlier is right—namely, that there is no alternative for one disposed to pursue the privileged observation route than to see the privilege as *necessarily* the exclusive property of the observing subject—then the incoherence of the Cartesian response is the incoherence of *any* broadly observational model of a subject's relation to their ordinary psychological states. That's the lesson I want to carry forward.

## IX

There is a proposal about our problem that for a time was widely accepted as Wittgenstein's own. In *Investigations* §308 he writes:

How does the philosophical problem about mental processes and states and about behaviourism arise?—The first step is the one that altogether escapes notice. We talk of processes and states and leave their nature undecided. Sometime perhaps we shall know more about them—we think. But that is just what commits us to a particular way of looking at the matter.

p. 34 And a little earlier (§304) he urged that we need to 'make a radical break with the idea that language always functions in one way, always serves the same purpose: to convey thoughts—which may be about houses, pains, good and evil, or anything else you please'. These sections advance the diagnosis that our difficulties in this neighbourhood are generated by 'the grammar which tries to force itself on us here' (§304). They go, Wittgenstein suggests, with a conception of avowals as *reports* and the associated conception of a self-standing subject-matter which they serve to report. We take it that there are mental states and processes going on anyway, as it were—the 'first step' that escapes notice—and that each person's avowals serve to report on such states and processes as pertain to her. The immediate effect is to set up a dilemma. How, in the most general terms, should we think of the states of affairs which confer truth on these 'reports'? There is the Cartesian—events-in-an-arena-accessible-only-to-the-subject—option; this does a neat job of explaining the distinctive marks of avowals, at least at a casual muster, but it relies on an 'analogy which . . . falls to pieces' (§308)—the analogy between avowals and observation reports made from a privileged position. But the only other option seems to be to 'go public': to opt for a view which identifies the truth-conferring states of affairs with items which are somehow wholly manifest and available to public view—an option which Wittgenstein expects, writing when he did, will naturally take a  $\hookrightarrow$  behaviourist shape, so that 'now it looks as if we had denied mental processes'. Of course, a philosopher who takes this option—whether in behaviourist or other form—will want to resist the suggestion that she is *denying* anything—according to *her* recommended understanding of 'mental process', just as Berkeley resisted the suggestion that he was denying the existence of matter. But the manifest problem is to reconcile any such conception of the truth-conditions of avowals with their distinctive marks: for as soon as you go public, it becomes obscure what advantage selves can enjoy over others.

This line of difficulty may seem to point to an obvious conclusion. Conceiving of avowals as reports of states and processes which are going on anyway appears to enforce a disjunction: *either* accept the Cartesian view, which cannot accommodate ordinary knowledge of others, *or* accept some form of externalization—perhaps behaviourist, nowadays more likely physicalist—which cannot sustain the special place of self-knowledge. So we should reject the parent assumption. And one tradition of commentary, encouraged especially by *Investigations* §244,<sup>24</sup> interprets Wittgenstein as doing this in a very radical way: as denying that avowals are so much as *assertions*—that they make statements, true or false—proposing to view them rather as *expressions* of the relevant aspects of the subject's psychology.<sup>25</sup>

Expression? To give expression to an aspect of one's psychology just means, presumably, to give it display, in the way in which wincing and a sharp intake of breath may display a stab of pain, or a smile may display that one is pleased, or a clenching of the teeth that one is angry. Propositional attitudes too can be open to natural expression of this kind: a prisoner's rattling the bars of his cell is a natural expression of a desire to get out. (It is not a way of acting on that desire, of course—it is not rationalized by it.) Wittgenstein's famous suggestion in §244 is that we should see the avowal of pain as an acquired form of pain behaviour—something one learns to use to supplant or augment the natural expression of pain and which, the *expressivist* tradition of commentary suggests, is no more a *statement*—something with a truth-evaluable content—than are such natural forms of expression.

p. 35 The immediate question is how well an expressivist treatment of avowals can handle their distinctive marks. And the answer appears to be: not badly at all. For instance, if the avowal 'I'm in pain' is not a statement, true or false, then naturally it is inappropriate to ask its author for grounds for it (Groundlessness), and naturally there is no question of her ignorance of its truth-value (Transparency). And if, when uttered with proper comprehension, it is to be compared to an episode of pain behaviour, then only its being a piece of dissimulation—not sincere—can stand in the way of a conclusion that the subject really is in pain (Strong Authority). (And of course it will provide a criterion for the subject's being in pain in just the way that ordinary pain behaviour does.)

Nevertheless, the expressivist proposal has come to be viewed more or less universally as a non-starter, for reasons preponderantly to do with the perceived impossibility of making coherent philosophy of language out of it. The claim that the avowal 'I am in pain' serves to make no statement, true or false, has to be reconciled with a whole host of linguistic phenomena whose natural explanation would exploit the opposed idea that it is, just as it seems, the affirmation of a truth-evaluable content. Here are four of the snags:

- (i) What has the expressivist proposal to say about transformations of tense—'I was in pain' and 'I will be in pain'? If either is a genuine assertion, doesn't there have to be such a thing as an author's making the *same* assertion at a time when doing so would demand its present-tense transform? If, on the other hand, they are regarded likewise merely as expressions, *what* do they serve to express? (Doesn't an expression have to take place at the same time as what it expresses?)
- (ii) How is the proposal to construe a locution like 'He knows that I am in pain'? If there is a use of the words 'I am in pain' so embedded, which I can use to express the content of someone else's possible knowledge, why may I not *assert* that very same content by the use of the same words?
- (iii) There are genuine—for instance, quantified—statements which stand in logical relations to 'I am in pain'. It entails, for instance, 'Someone is in pain'. How can a genuine statement be entailed by a mere expression?
- (iv) 'I am in pain' embeds like any normal assertoric content in logical constructions such as negation and the conditional. 'It's not the case that I am in pain' and 'If I am in pain, I'd better take an aspirin' are syntactically perfectly acceptable constructions. But how can a mere expression, contrast: assertion, be

*denied?* And doesn't the antecedent of a conditional have to be understood as the hypothesis that *something is the case?*

p. 36 This kind of point—I shall dub the whole gamut 'the Geach point'<sup>26</sup>—has often been used as a counter to various forms of expressivism, notably in ethics, and much ingenuity has been expended (squandered?) by philosophers of expressivist inclination in the attempt to meet it. But in the present case I don't think it ought to have been influential at all. In the ethical case, the expressivist thesis is, crudely, that there are *no* real moral states of affairs; so the occurrence of what are apparently truth-evaluable contents couched in distinctively moral vocabulary has to be some kind of illusion. In that case the Geach point represents a very serious challenge, since it seems to show that everyday moral thought, in exploiting perfectly standard syntactic resources like those afforded by ordinary sentential logic, requires, to the contrary, that truth-evaluable moral contents exist. By contrast, it is no part of the present, allegedly Wittgensteinian expressivist proposal that there is no such thing as a statement of ordinary psychological fact. No one is questioning that 'He is in pain' is an assertion. The expressivist thesis distinctively concerns *avowals*.

How does that difference help? Well, it is clear that we have to draw a distinction in any case between the question whether an indicative sentence is associated with a truth-evaluable content and the question whether its characteristic use is actually *assertoric*. For the two notions routinely come apart in the case of standard performatives like 'I promise to be on time', 'With this ring, I thee wed', 'I name this ship . . .', and so on. Each of these locutions embeds in all the ways the generalized Geach point focuses on; and none of them is standardly used, in the atomic case, as an assertion. We should conclude that what the Geach point signals is merely the presence of truth-evaluable content. It is powerless to determine that the standard use of a locution is to *assert* such a content. And now the expressivist thesis about avowals can be merely that the typical use of such sentences is as expressions rather than assertions. There need be no suggestion that one *cannot* make assertions about one's own psychology. But the suggestion—now initially rather exciting—will be that the appearance of the *epistemic* superiority of the self which avowals convey is an illusion created by attempting to find a home for features of such utterances which they carry *qua* expressions in the context of the mistaken assumption that they are ordinary assertions. When selves *do* make strict assertions about their own

p. 37 psychology, the story should continue, any epistemic advantages ↵ they enjoy are confined to those of superiority of evidence which I noted briefly at the beginning.

That, it seems to me, is, in outline, how the best expressivist proposal should go. Now to its real problems. Perhaps the most immediate awkwardness, if a general account of avowals is to be based upon the §244 idea, is that, even in the case of sensations, the range of cases where there are indeed *natural*, non-linguistic forms of expression—cases like pains, itches, and tickles—is very restricted: contrast, for instance, the sensation of coolness in one foot, or the smell of vanilla. In the latter kind of case, the suggested model of the acquisition of competence in the avowal simply won't grip, and the theorist will have to try to live with the idea of a range of sensations whose *only* expression consists in their avowal. The same is evidently true in Spades of psychological items other than sensations. This threatens a worrying dilution of the key notion of *expression*.

That's a worry that might, I suppose, be worked on. But the next one seems decisive. Suppose a highly trained secret agent under torture resolutely gives no ordinary behavioural sign of pain. However, his torturers are men of discernment, with subtle instruments, who know full well of his agony, none the less: they know the characteristic signs—patterns on the electroencephalograph, raised heart rate, activation of reflexes in the eye, changes in surface skin chemistry, etc. If the suggestion really is to be that the superiority of the first-person viewpoint is *wholly* an artefact of a grammatical misunderstanding—the misconstrual of expressions as assertions—then, any *knowledge*, strictly so conceived, which the victim has of his own pain has to originate in the same way as that of his tormentors. But by hypothesis he isn't expressing pain behaviourally. And the signs that leave them in no doubt are things which, in his agony, he may not be attending to, or which, like the print-out on the electroencephalograph, he may not be able to see, or interpret if he could see. So in such a case, when

it comes down to knowledge, it looks as though the expressivist account must represent the victim as actually *at a disadvantage*. That's evident nonsense.

In general, merely to conceive of avowals as expressive does not, when it goes in tandem with an acceptance of the reality of the states of affairs which they express, provide any way of deflecting the question: how, broadly speaking, should we conceive of the kind of state of affairs which is apt to confer truth on psychological ascriptions, and in what sort of epistemological relationship do their subjects themselves in general stand to such states of affairs? If this relationship is in any way more than evidentially privileged, we have our original problem back. If it isn't, we seem to get absurdities like that just illustrated.

p. 38 A different way of seeing the unplayability of the expressivist position is to reflect that the content of an avowal is always available to figure just in a subject's thoughts, without public expression. You may sit reading and think to yourself, 'My headache has gone', without giving any outward sign at all. And anyone versed in ordinary psychology will accept that *if* you have that thought, not by way of merely entertaining it, but as something you endorse, then you will be right (Authority); that there is no way that your headache could have passed unless you are willing to endorse such a thought (Transparency); and that your willingness to endorse it will not be the product of inference or independently formulable grounds (Groundlessness). Thus analogues of each of the marks of avowals that pose our problem engage the corresponding unarticulated thoughts. It must follow that the correct explanation of the possession of them by avowals cannot have anything to do with illocutionary distinctions.

We should conclude that the expressivist proposal flies rather further than is usually thought. But it is a dead duck all the same.

## X

For sure, the textual evidence for attributing the expressivist view to Wittgenstein was always pretty exiguous. *Investigations* §244 in particular should be contrasted with the much more cautionary and nuanced remarks elsewhere.<sup>27</sup> Such apparent equivocations, of course, are fuel for the common complaint that while Wittgenstein has suggestive criticisms to offer of certain tendencies in the philosophy of mind, he left any intended positive contribution shrouded in fog. What exactly—or even roughly—is Wittgenstein saying about avowals, if he is not proposing the expressivist view? How exactly does he propose we should liberate our thinking from Cartesian tendencies? What did he think we should put in their place?

p. 39 Well actually, I don't think it is all that difficult to glean what his positive recommendation is, at least in general outline. The difficulty is, rather, to live with it. The first essential in interpreting him here is to give due prominence to the *Investigations'* explicit conception of the genesis of philosophical problems and of proper philosophical method. Wittgenstein wrote, recall, that ↵

we may not advance any kind of theory. There must not be anything hypothetical in our considerations. We must do away with all *explanation*, and description alone must take its place. . . . [Philosophical problems] are solved, rather, by looking into the workings of our language, and that in such a way as to make us recognize those workings: *in despite of* an urge to misunderstand them. . . . Philosophy is a battle against the bewitchment of our intelligence by means of language.<sup>28</sup>

And, very famously: 'Our mistake is to look for an explanation where we ought to look at what happens as a "proto-phenomenon". That is, where we ought to have said: *this language-game is played*.'<sup>29</sup>

The bearing of these strategic remarks is immediate if we reflect that our whole problem is constituted by a demand for explanation. We are asking: what is the *explanation* of the characteristic marks of avowals? And we

easily accept a refinement of the question along the lines: what is it about the subject-matter of avowals, and about their authors' relation to it, which explains the possession by these utterances of their characteristic effortless, non-inferential authority? Cartesianism takes the question head on, giving the obvious, but impossible, answer. And the expressivist proposal, radical though it is in its questioning of the assumption that the authority of an avowal is the authority of a claim to truth, is not so radical as to raise a question about the validity of the *entire explanatory project*. But Wittgenstein, seemingly, means to do just that. Against the craving for explanation, he seemingly wants to set a conception of the 'autonomy of grammar'.<sup>30</sup> The features of avowals which set our problem—the features which seem to betray something remarkable about self-knowledge—do so only if we suppose that they are in some way *consequential* upon something deeper: for instance, the nature of their subject-matter and of their author's relationship to it. But what imposes that way of looking at the matter? Why shouldn't psychological discourse's exhibition of these features be regarded as primitively constitutive of its being *psychological*, so that the first-third-person asymmetries that pose our question belong primitively to the 'grammar' of the language-game of ordinary psychology, in Wittgenstein's special sense—'grammar' which 'is not accountable to any reality' and whose rules 'cannot be justified by showing that their application makes a representation agree with reality'?<sup>31</sup>

p. 40 What did Wittgenstein suppose entitled him to this? In his later work, as everyone knows, he radically rethought his early conception of the relation between language and reality. It is to this readjustment, I suggest, that we must look if we are to understand the doctrine of the 'autonomy of grammar'. As I read the early 300s, the obstacle which Wittgenstein sees as lying in the way of our philosophical understanding of 'mental processes and states' is not the assumption of the truth-evaluability of avowals, as the expressivist interpretation has it, but rather a general picture of the working of *truth-evaluable* language. Wittgenstein means to reject a certain picture of what truth-evaluability involves: the picture gestured at in §304, that our statements always serve 'the same purpose: to convey thoughts—which may be about houses, pains, good and evil, or anything else you please'. This picture involves thinking of assertions as expressing propositions which are laid over against reality in the manner of the *Tractatus*, so that there have to be self-standing states of affairs to correspond to avowals, when they are true, and it has therefore to be possible to raise general questions about the nature of these self-standing states of affairs, and the nature of the subject's knowledge of them. And then, when we are mindful of the distinctive marks of avowals, it appears that the states, and the mode of knowledge, must be something rather out of the ordinary—the relevant states of affairs have to be conceived as somehow especially transparent to the subject, or, at the least, as working on her by some form of curiously reliable 'blindsight'—whose curious reliability, moreover, would have to be common knowledge if the authority credited to avowals is to be explained. Wittgenstein's diagnosis is that the 'philosophical problem about mental processes and states and about behaviourism' arises because we insist on interpreting the truth-evaluability of avowals—the source of the linguistic features on which the Geach point fastens—as imposing a conception of their being true, when they are, in terms which have to raise these constitutive questions about nature and access. But these are the very questions, Wittgenstein is saying, which we must free ourselves from the temptation to raise; they are the questions which lead to the fast track into the fly-bottle.

Of course, the conception of truth and truth-makers which, in Wittgenstein's diagnosis, is here at the root of our difficulty is the core of the outlook which Hilary Putnam has called 'metaphysical realism'. Perhaps the single most significant departure effected in Wittgenstein's later philosophy is his coming to believe that we have to stop thinking about the relationship between language and reality, and about the truth-predicate, in that kind of way.

Wittgenstein's stance deserves a much finer-grained depiction than I can attempt here. If we abstract from the globally anti-explanatory background mantra, the cash value of the proposal, just for the issue of self-knowledge, involves a generalization to all avowable subject-matter, phenomenal and attitudinal, of a view which might be characterized like this:

the authority standardly granted to a subject's own beliefs, or expressed avowals, about his intentional states is a *constitutive principle*: something which is not a consequence of the nature of those states, and an associated epistemologically privileged relation in which the subject stands to them, but enters primitively into the conditions of identification of what a subject believes, hopes and intends.<sup>32</sup>

I'll call this general viewpoint the 'Default View'. According to the Default View, it is just primitively constitutive of the acceptability of psychological claims that, save in cases whose justification would involve active self-interpretation, a subject's opinions about herself are default-authoritative and default-limitative: unless you can show how to make better sense of her by overriding or going beyond it, her active self-conception, as manifest in what she is willing to avow, must be deferred to. The truth-conditions of psychological ascriptions are primitively conditioned by this constraint. In particular, it is simply basic to the competent ascription of the attitudes that, in the absence of good reason to the contrary, one must accord correctness to what a subject is willing to avow, and limit one's ascriptions to her to those she is willing to avow.

It would be a great achievement of Wittgenstein's discussion if it made it possible to understand how the Default View need not be merely an unphilosophical turning of the back. But it is anything but clear, actually, how a repudiation of the metaphysical realist picture of truth could just by itself directly enjoin this conception. Moreover, it is very difficult to rest easy with the general anti-explanatory mantra. For it is seemingly in tension with a diagnostic thought which is very important to Wittgenstein himself: that philosophical problems characteristically arise because we are encouraged by surface-grammatical analogies to form expectations about an area of discourse which are appropriate only for a particularly salient surface-grammatical analogue of it. That is *exactly* Wittgenstein's diagnosis in the present case: the target analogy is that between the use of avowals and ordinary reports of observation. ↪ So then that diagnosis itself requires that the explanatory questions which we are required *not* to press in the case of avowals are, by contrast, perfectly properly raised, and answerable, in the case of ordinary reports of observation. There cannot, accordingly, just be a blanket prohibition against explanatory questions of that kind. Put that simple thought alongside the plausible claim that there are perfectly legitimate modes of *conceptual* explanation—informal mathematics, in particular, is full of them—and it appears that it cannot in general be merely a confusion to seek to explain features of the practice of a discourse a priori by reference to our concepts of the kind of subject-matter it has and of the epistemic capacities of speakers. Thus the insistence that these questions are misplaced in the target case of psychological self-ascriptions begins to seem merely dogmatic.

Is there any way this impression of dogmatism might be dispelled? In the analogy of the kaleidoscope, our conception is that of a range of independent features and events—evolving patterns of shape and colour—to which the privileged observer is sensitive—responsive—by dint of his situation and his possession of certain germane cognitive capacities, notably vision. There is a story to be told about the kind of things on display and how things of that kind can elicit a response from someone with a suitable cognitive endowment. Now one way to try to exculpate the Default View from the charge of dogmatism, it seems to me, is to seek a framework which places controls on the relevant idea of *responsiveness*.<sup>33</sup> One form of control might be elicited from pursuing certain recently much discussed issues to do with judgement dependence and the *Euthyphro contrast*: we may pursue the details of the relations, in different regions of thought, between best opinion and truth, attempting thereby to arrive at a conception of what it is for them to relate too closely, so to speak, for their

p. 43

congruence to count as a *success in tracking*. Another control might emerge from consideration of the question of how wide the potential *explanatory range* has to be of a certain type of state of affairs if we are to think of our judgements about it as genuinely responsive to its subject-matter at all (*Width of Cosmological Role*).<sup>34</sup> We can seek a framework of such controls and try to show that first-person psychological discourse emerges on the wrong side of the tracks, so to speak, under the application of such controls. Then, if its apparent urgency does indeed derive from a tacit assumption of the *responsiveness* of selves to their own psychological states, the general explanatory question about self-knowledge, which official Wittgensteinian philosophical method would have us ignore, can emerge as something which we can understand *why* we ought not to ask.

By contrast, lacking such a framework, and the right kind of upshot within it concerning psychological discourse, it is hard to see how the Default View can come to much more than a take-it-or-leave-it recommendation: a mere invitation to *choose* to treat as primitive something which we have run into trouble trying to explain, and to do so just on that account. Wittgenstein notoriously came to view philosophical problems as akin to a kind of self-inflicted intellectual *disease*; they would thus contrast starkly with *mathematical* problems as traditionally viewed (not by Wittgenstein, of course)—a kind of sublime, objective puzzle whose force can be felt by any rational intellect. If philosophical problems are justly deflated in Wittgenstein's way, then a kind of 'Here: think of matters this way, and you'll feel better' remedy might be the best we can do. But the prospect immensely disappoints. For most of us, after all, the attraction of philosophy is all about gaining understanding. Except in cases where one can *explain* a priori why the demand for understanding is inappropriate, it is apt to seem like a mere abrogation of the subject to be told one mustn't make it at all.

## XII

We owe to Wittgenstein the insight that we are making an assumption in regarding it as a deficiency of understanding to lack a satisfactory explanation of the distinctive marks of avowals. The assumption is, roughly, that those distinctive marks must be *consequential*: that they must either derive from the nature of the subject-matter—something which therefore drives our discourse about it into the relevant characteristic turns—or else they must derive from some unobvious feature of the semantics of first-person psychological discourse (its being, for instance, expressive rather than assertoric). So, according to the assumption, there must be an explanation which we have yet to assemble and get into focus.

p. 44

There is a frontal collision between this way of thinking and the conception of the nature of legitimate philosophical enquiry seemingly quite explicit in Wittgenstein's later official methodological pronouncements. According to Wittgenstein, the limit of our philosophical ambition should be to recognize the assumptions we are making in falling into philosophical difficulty, and to see our way clear to accepting, by whatever means, that nothing forces us to make them. It is, for Wittgenstein, the very craving for legitimizing explanations of features of our talk about mind, or rules, or mathematics that leads us into hopeless puzzles about the status—the epistemology and ontology—of those discourses. Philosophical treatment is wanted not to solve these puzzles but to undermine them—to assuage the original craving that leads to the construction of the bogus models and interpretations by which we attempt to make sense of what we do, but which are the source of all our difficulties, and yet whose want is felt as a lack of understanding. The problem of self-knowledge is a signal example. It can have—I believe Wittgenstein thought—no solution of the kind we seek; for that very conception of a solution implicitly presupposes that there must be a something-in-virtue-of-which the distinctive marks of avowals are sustained. But those marks are part of 'grammar', and grammar is not sustained by anything. We should just say 'This language-game is played.'

The generalization of this position—the execration of all philosophical explanation—seems to me vulnerable to a version of what one might call 'the paradox of postmodernism'. The paradox is that while, like all

'deflationists', Wittgenstein needs to impress us of the illegitimacy of more traditional aspirations, *argument* for that is hard to foresee if it is not of the very coin which he is declaring to be counterfeit. For what is needed here is precisely a *philosophical explanation*. Sure, what belongs to 'grammar', in Wittgenstein's special sense of that term, requires no explanation. *Of course*; that's a matter of definition. But even a sympathetic reading of him will find a frustrating inattention to the question of when something may legitimately be taken to be part of 'grammar'. It may be a crucial first step to recognize that the problem of self-knowledge is occasioned by an assumption of explicability—an assumption that may be discarded with a clear conscience if the special position of subjects in determining what is true of their psychology is indeed 'grammatical'. But, one wants to say, what shows that? Once one recognizes the Default View as a possibility, the immediate instinct is to ask: what might *justify* the idea that it is the whole truth? That is the instinct to attempt to understand when and why it is a good move to dismiss the attempt to understand. To succumb is to re-enter the space of explanatory philosophy. To resist is to have no reason for the Default View.

p. 45 To feel this dissatisfaction is not to have a reason to deny the insight that in a wide class of cases philosophical perplexity does indeed take the form of a casting about for what strike us as satisfying explanations of features of our language, and of failing to find any that do not generate singularities of one sort or another. (Just briefly to mention a second prominently Wittgensteinian example: how are we to make sense ↵ of the intelligibility of the distinction between whether a statement is really true and whether anybody ever takes it to be true unless the rule incorporated in its truth-condition may be thought of as issuing its verdict autonomously and independently of any human judgement? So isn't the very idea of unratified truth an implicit commitment to 'rules-as-rails' platonism?!) But to accept Wittgenstein's insight, that some of the hardest-seeming philosophical problems take this form, is not a commitment to an explanation-proscribing view of philosophy. Even if it is misguided to persist in assuming that there must be *something* that can satisfactorily take up the explanatory slack left by the demise of platonism, or Cartesianism, it may yet be possible to explain why such an assumption need not be true in particular cases. It does not seem merely confused to seek, in particular, to characterize with some care the conception we have of the kinds of ways the marks of avowals might in principle be explained. It is even foreseeable that such a characterization might lead to a clear-headed realization that nothing could fulfil it. *That* would be the discovery that, in this area, 'gives philosophy peace' (§133). ↵

p. 46

## Notes

- 1 Jane Austen, *Emma*; Penguin Books edn. (London, 1987), 398. I borrow this nice example from Julia Tanney's 'A Constructivist Picture of Self-Knowledge', *Philosophy*, 71 (1996), 405–22 [10.1017/S0031819100041668](https://doi.org/10.1017/S0031819100041668)<sup>↗</sup>.
- 2 I am indebted in this section to many discussions, years ago, with Andy Hamilton.
- 3 Ludwig Wittgenstein, *The Blue and Brown Books* (Oxford: Blackwell, 1964), 66–7.
- 4 The term is originally Shoemaker's, I believe. It is taken up by Gareth Evans in *The Varieties of Reference* (Oxford: Oxford University Press, 1982); see esp. ch. 7, sect. 2.
- 5 See his 'Self-Reference and Self-Awareness', *Journal of Philosophy*, 65 (1968), 568 ff.
- 6 Wittgenstein, *Blue and Brown Books*, 67.
- 7 See e.g. Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953), §§34, 146, 152, 154, 205, 303, 330–2, 427, 577, 673; also part II, §vi, p. 181, and §xi, pp. 217–18. The distinction is prominent in the *Remarks on the Philosophy of Psychology* (trans. G. E. M. Anscombe (Oxford: Blackwell, 1980)) as well, where Wittgenstein uses the terminology of *dispositions* versus *states of consciousness*; see e.g. vol. ii, §§45, 48, 57, and 178.
- 8 No speedy resolution is to be expected, for Wittgenstein was content to gesture at his distinction quite loosely—perhaps believing that it is vague, that the concept of a 'mental process' is a family resemblance concept, or whatever.
- 9 John McDowell, 'Intentionality and Interiority in Wittgenstein', in K. Puhl (ed.), *Meaning Scepticism* (Berlin: de Gruyter, 1991), 148–69, at 149.
- 10 It is, in a sense, the entire subject-matter of George Eliot's novella, *The Lifted Veil*.
- 11 In particular, I do not think that we have any satisfactory concept of what it would be to be in touch with others' mental

- states *telepathically*. I do not mean, of course, to rule it out that someone might prove, by dint of *his own* occurrent suspicions and afflictions, to be a reliable guide to the states of mind of another. But that possibility falls conspicuously short of the idea that a subject might share direct witness of another's mental states.
- 12 See n. 7.
- 13 See Crispin Wright, 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention', in Puhl (ed.), *Meaning Scepticism*, 126–47.
- 14 McDowell, 'Intentionality and Interiority in Wittgenstein'.
- 15 Ibid. 164–5.
- 16 Wittgenstein, *Philosophical Investigations*, §6.
- 17 Wright, 'Wittgenstein's Later Philosophy of Mind', 146.
- 18 McDowell, 'Intentionality and Interiority in Wittgenstein', 163–4.
- 19 This idea is elaborated a little in my 'Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics', in Alexander George (ed.), *Reflections on Chomsky* (Oxford: Blackwell, 1989), 233–64, at 237 ff.
- 20 Wittgenstein, *Philosophical Investigations*, pt. II, §iii.
- 21 McDowell, 'Intentionality and Interiority in Wittgenstein', 152.
- 22 See his 'One Strand in the Private Language Argument', *Grazer Philosophische Studien*, 33/4 (1989), 287–303. This paper should now be taken in the context of the more general ideas about the interface between thought and the external world presented in McDowell's John Locke lectures, *Mind and World* (Cambridge, Mass.: Harvard University Press, 1994).
- 23 McDowell, 'Intentionality and Interiority in Wittgenstein', 158.
- 24 But see also *Remarks on the Philosophy of Psychology*, vol. i, §§450, 501, 593, 599, and 832.
- 25 The sometime popularity of this interpretation is traceable to its being advanced by several of the first reviewers: P. F. Strawson, for instance, in his critical study of the *Investigations* in *Mind*, 53 (1954), 70–99; and Norman Malcolm in his 'Wittgenstein's *Philosophical Investigations*', *Philosophical Review*, 53 (1954), 530–59.
- 26 After P. T. Geach's emphasis of such difficulties for moral expressivism, Austin's performatory account of knowledge, etc. See his 'Assertion', *Philosophical Review*, 74 (1965), 449–65 [10.2307/2183123](https://doi.org/10.2307/2183123)<sup>24</sup>.
- 27 e.g. *Investigations*, pt. II, §ix: 'a cry, which cannot be called a description, which is more primitive than any description, for all that serves as a description of the inner life. A cry is not a description. But there are transitions. And the words "I am afraid" may approximate more, or less, to being a cry. They may come quite close to this and also be *far* removed from it.'
- 28 Ibid. §109.
- 29 Ibid. §654. It doesn't matter that this is said in the context of discussion of a different issue (recollection of the content of a prior intention).
- 30 As Baker and Hacker style it.
- 31 Wittgenstein, *Philosophische Grammatik*, ed. R. Rhees, tr. A. Kenny (Oxford: Blackwell, 1974), sect. X, §§133 and 134.
- 32 Wright, 'Wittgenstein's Later Philosophy of Mind', 142.
- 33 I suppose this is a programme of what McDowell has disparagingly called 'constructive philosophy'.
- 34 Both these ideas are explored in my *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press, 1994).