

# Royal Institute of Philosophy Supplement

<http://journals.cambridge.org/PHS>

Additional services for *Royal Institute of  
Philosophy Supplement*:

Email alerts: [Click here](#)

Subscriptions: [Click here](#)

Commercial reprints: [Click here](#)

Terms of use : [Click here](#)



---

## Self-knowledge: the Wittgensteinian Legacy

Crispin Wright

Royal Institute of Philosophy Supplement / Volume 43 / March 1998, pp 101 - 122  
DOI: 10.1017/S135824610000432X, Published online: 08 January 2010

**Link to this article:** [http://journals.cambridge.org/abstract\\_S135824610000432X](http://journals.cambridge.org/abstract_S135824610000432X)

### How to cite this article:

Crispin Wright (1998). Self-knowledge: the Wittgensteinian Legacy. Royal Institute of Philosophy Supplement, 43, pp 101-122 doi:10.1017/S135824610000432X

**Request Permissions :** [Click here](#)

# Self-knowledge: the Wittgensteinian Legacy

CRISPIN WRIGHT

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.

In this paper<sup>1</sup> I'll begin by trying to characterize this dominion with some care. We need to have it as clear as possible why one traditional way of thinking about the matter has seemed so attractive – even unavoidable – and what a satisfactory account of the issues in this region has to accomplish. However my overarching concern will be with the bearing of later Wittgensteinian materials on the question. Ultimately I think we can get an insight into the intended force of something which I do not think has so far been sufficiently 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 nonetheless a truism that for the most part we know ourselves best – better than we know others and better than they know us.

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

<sup>1</sup> This is an edited version of my contribution to *Knowing Our Own Minds*, ed. C. Macdonald, B. Smith and C. Wright (Oxford: Oxford University Press 1998).

## Crispin Wright

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 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* – comprises 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 produces 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.

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 its author. Since we standardly credit any interlocutor, absent 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

## Self-knowledge: the Wittgensteinian Legacy

accepted by others, on pain 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 they contain – 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 when 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>2</sup>

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 dependable, somewhat avuncular part of the background to her life of whom she

<sup>2</sup> Jane Austen, *Emma* (London: Penguin Books, 1987), p. 398. This nice passage was drawn to my attention by Julia Tannev, who uses it for her own purposes in her 'A Constructivist Picture of Self-Knowledge', *Philosophy* 71 (1996), 405–22.

## Crispin Wright

has entertained no romantic notions. 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 first as something she 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 no way incongruous if the subject professes ignorance of particular aspects of her intentional psychology. However, 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.

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 the matter is not basic but calls for

## Self-knowledge: the Wittgensteinian Legacy

(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 other explanations of a subject's willingness to assert a bogus avowal than those provided by misunderstanding, insincerity or misinterpretation. This is indeed 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.

It is striking that attitudinal avowals would appear to exhibit a form of weak authority nevertheless: that is, at least in basic, non-self-interpretative cases they provide empirically assumptionless justification for the corresponding third-person claims.<sup>3</sup> Other things being equal, I *ought to know* what my beliefs, desires and hopes, etc., are, even if sincerity and understanding alone do not guarantee the truth of what I say about them.

<sup>3</sup> Since it cannot be attributed, as with phenomenal avowals, to the fact of sincerity-cum-understanding guaranteeing truth, it is an interesting question what this weak authority should be taken to 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 and 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

We now have a sufficient focus for our central question. The cardinal problem of self-knowledge is that of explaining *why* avowals display the marks they do – what is it 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, transparency and strong authority in the case of phenomenal avowals, and with groundlessness, transparency and weak authority 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?

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 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:

1. The demand that he produces reasons or corroborating evidence for his claims will be misplaced – the most he will be able to say is that he is the only one in a position to see, and that is how things strike him;

2. 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 that 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

3. where P is any claim about the patterns of shape and colour visible within, there will be no provision – bearing in mind Hero’s assumed perceptual competence and attentiveness – for his intelligibly professing ignorance whether or not P.

This 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 others’ avowals. But once into one’s stride with this type of thinking, this difference will not seem

## Self-knowledge: the Wittgensteinian Legacy

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. That 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>4</sup> McDowell recoils against 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>5</sup> It is arguably a presupposition of the whole idea of the continuation of one's consciousness after death. The thought

<sup>4</sup> From p. 149 of John McDowell, 'Intentionality and Interiority in Wittgenstein', in *Meaning Scepticism*, ed. K. Puhl (Berlin: de Gruyter, 1991), pp. 148–69.

<sup>5</sup> To take another nineteenth-century example, it is, in a sense, the entire subject matter of George Eliot's novella, *The Lifted Veil*. The Cartesian character of that writer's notions about the mental is explored in depth in Catherine Wright's 'The Unseen Window: Middlemarch, Mind and Morality' (Ph.D. dissertation, University of St Andrews, 1991).

## Crispin Wright

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, very simply, 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 (anyway, none of which we have any remotely satisfactory concept<sup>6</sup>) whose suspension would put other ordinary people in position to avow away on my behalf, as it were – would transfer, or extend my advantage to them. 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.

### III

If this is right, then a deconstruction of the privileged-observation solution to the problem of self-knowledge is an indispensable prerequisite of an overall satisfactory philosophy of mind. It seems to me that the accomplishment of such a deconstruction was one major achievement of Wittgenstein's later philosophy – though it would take another paper, or series of papers, properly to fill out how it

<sup>6</sup> 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.

## Self-knowledge: the Wittgensteinian Legacy

goes.<sup>7</sup> In essentials, what he does is to mount a two-pronged attack on the Cartesian picture, with the prongs corresponding to the distinction between the two main kinds of avowals. The idea that phenomenal avowals serve as inner observation reports is challenged by the so-called 'private language argument' – the battery of considerations that surface in §§243 to the early 300s in the *Investigations*. The attack is multi-faceted but the famous central strand is that the Cartesian picture implicitly surrenders the resources needed for a distinction which is essential if such '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, which demands a potential contrast between how matters seem to an observer and how they really stand. The corresponding conception of attitudinal avowals, by contrast, 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>8</sup> A central problem with the idea that attitudinal avowals describe introspectable mental occurrences 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 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 conception of attitudinal avowals as reports of inner observation thus stands at odds with another, fundamental feature of their grammar – their essential answerability, in broadly the fashion of dispositions, to matters which may be unobservable at the time they are asserted.<sup>9</sup>

<sup>7</sup> For further indications, see my 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention', in *Meaning Scepticism*, ed. Puhl, pp. 126–47.

<sup>8</sup> See e.g. *Philosophical Investigations* §§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* as well, where Wittgenstein uses the terminology of *dispositions* versus *states of consciousness*; see, for instance, vol. II, §§45, 48, 57 and 178.

<sup>9</sup> This idea is elaborated a little at pp. 237ff. Of my 'Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics', in *Reflections on Chomsky*, ed. Alexander George (Oxford: Basil Blackwell, 1989), pp. 233–64.

## Crispin Wright

The pursuit of these ideas of Wittgenstein leads one to recognize deep incoherences in the Cartesian response to our problem – incoherences that are prior to its inordinate sceptical costs. Note, moreover, that if what I said earlier is right – viz. 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 her ordinary psychological states. That's a crucial lesson.

### IV

But if not an observational model, then what? There is a proposal about our problem that for a time was widely accepted as Wittgenstein's own. At *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.

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) –

## Self-knowledge: the Wittgensteinian Legacy

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 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 advantages 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>10</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>11</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

<sup>10</sup> But see also *Remarks on the Philosophy of Psychology*, vol. I, §§450, 501, 593, 599 and 832.

<sup>11</sup> 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* 63 (1954), 70–99; and Norman Malcolm in his 'Wittgenstein's *Philosophical Investigations*' in *The Philosophical Review* 63 (1954), 530–59.

## Crispin Wright

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.

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 am 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 more or less universally viewed 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:

1. 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 accompany – hence have to take place at the same time as – what it expresses?)

2. 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?

3. 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?

4. ‘I am in pain’ embeds like any normal assertoric content in logical constructions such as negation and the conditional. ‘It’s not the

## Self-knowledge: the Wittgensteinian Legacy

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

This kind of point – I shall dub the whole gamut 'the Geach point'<sup>12</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 how 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

<sup>12</sup> After P. T. Geach's emphasis of such difficulties for moral expressivism, Austin's performatory account of knowledge, etc. See Geach's 'Assertion' in *The Philosophical Review* 74 (1965), 449–65.

## Crispin Wright

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 psychology, the story should continue, any epistemic advantages they enjoy are confined to those of superiority of evidence which I briefly noted 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 nonetheless: they know the characteristic signs – patterns on the electro-encephalograph, raised heart rate, activation of reflexes in the eye, changes in surface skin chemistry, etc., 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 electro-encephalograph, 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

## Self-knowledge: the Wittgensteinian Legacy

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.

A different way of seeing the ultimate 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 while the expressivist proposal flies rather further than is usually thought, it is a dead duck all the same.

## V

For sure, the textural 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>13</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 advancing 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?

<sup>13</sup> For instance *Investigations* part II, section ix: 'a cry, which cannot be called a description, which is more primitive than any description, for all that it 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.'

## Crispin Wright

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 settle for 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 recognise 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>14</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>15</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 such lines as: 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 characteristically 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>16</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

<sup>14</sup> *Philosophical Investigations* §109.

<sup>15</sup> *Philosophical Investigations* §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).

<sup>16</sup> As Baker and Hacker style it.

## Self-knowledge: the Wittgensteinian Legacy

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>17</sup>

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 *all* 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

<sup>17</sup> *Philosophische Grammatik*, (Oxford: Blackwell, 1969), section X, §§133 and 134.

## Crispin Wright

– which we must free ourselves of 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 his single most significant departure from the metaphysics of the *Tractatus* was Wittgenstein's coming to believe that we have to stop thinking about the relationship between language and reality, and about truth, in that kind of way.

## VI

What is involved in this re-orientation deserves a more refined depiction that 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>18</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, absent 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 might be the

<sup>18</sup> From p. 142 of my 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention', in *Meaning Scepticism*, ed. Puhl.

## Self-knowledge: the Wittgensteinian Legacy

last word on the issue. 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 difficult to rest easy with the general anti-explanatory mantra, which 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 unanswerable, in the case of ordinary reports of observation. There cannot, accordingly, just be a blanket prohibition against explanatory questions of that kind. Put that 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>19</sup> One form of control might be elicited from pursuing recently prominent 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 congruence to count as a *success in tracking*.

<sup>19</sup> I suppose this is a programme of what McDowell has disparagingly called 'constructive philosophy'.

## Crispin Wright

Another control might emerge from consideration of the question how wide the potential *explanatory range* has to be of a certain type of states of affairs if we are to think of our judgements about them as genuinely responsive to their subject matter at all (*Width of Cosmological Role*).<sup>20</sup> We can seek a general 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 their application. 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, might emerge as something which we can understand *why* we ought not to ask.

By contrast, if that kind of project is dismissed, 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 remedy seems enormously disappointing, intellectually. 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 quest is inappropriate, it is apt to seem a mere abrogation of the subject to be told there is nothing to understand.

## VII

Let me try to draw some strands together. 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 either they must 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 seman-

<sup>20</sup> Both these ideas are explored in my *Truth and Objectivity* (Cambridge, MA: Harvard University Press, 1994).

## Self-knowledge: the Wittgensteinian Legacy

tics 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.

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, with the very craving for *legitimizing explanations* of features of our talk about mind, or rules, or mathematics, that we are led 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 'estoppel' 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*. To be 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 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

## Crispin Wright

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.

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 unrati ed 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* satisfactorily to take up the explanatory slack left by the demise of platonism, or Cartesianism, it may yet be possible to explain why such an assumption needn't 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 – at least in this area – the discovery 'that gives philosophy peace'.<sup>21</sup>

<sup>21</sup> The phrase, of course, is from *Philosophical Investigations* §133.