

The Simplicity of the Soul

Jonathan Bennett

From: *Journal of Philosophy* 64 (1967), pp. 648–660.

1

The second Paralogism in Kant's *Critique of Pure Reason* argues that the soul is noncomposite, or lacks parts, because 'the action of [the soul] can never be regarded as the concurrence of several things acting' (A 351). Although not an obviously nourishing philosophical topic, this is less jejune than it looks. Kant argues that 'The soul is noncomposite' (S) when properly understood does express an a priori truth; that the latter is 'empty' in a special way; and that it is peculiarly liable to be misunderstood as having content. Each part of his argument raises matters worth discussing.

Here is Kant's reason for thinking that S says something true. If something is to be viewed as composite, some mind must unite its constituent parts by an 'intellectual synthesis'—must combine them in the judgment that they somehow go together. My basic notion of compositeness is that of a number of elements united by my judgment: it pre-requires myself, my synthesizing self, to judge that the elements are so interrelated as to *compose* a whole. I therefore cannot bring this notion to bear upon *myself*:

Although the whole of the thought could be divided and distributed among many subjects, the subjective 'I' can never be thus divided and distributed, and it is

this 'I' that we presuppose in all thinking (A 354).

This argument restricts 'What is the soul like?' to 'What can I regard my soul as being like?' Before discussing that vital restriction, let us see how Kant operates within it.

An ordinary statement describing something as noncomposite has implications of the form: 'If... my experience will be thus and not so', e.g. by implying that if the thing is hammered it won't split. Kant insists that the truth expressed by S has no such implications:

It tells us nothing whatsoever in regard to myself as an object of experience. . . It concerns only the condition of our knowledge; it does not apply to any assignable object (A 356).

S gets its truth from the fact that I can be aware of a composite only if I stand in judgment over all its parts. This, however, is a fact about 'the condition of [my] knowledge,' not about its content:

The judgment 'I think' . . . is the vehicle of all concepts. . . It can have no special designation, because it serves only to introduce all our thought, as belonging to consciousness (A 341).

Kant's view about S, then, is grounded in something more

general: my inability to apply 'composite' to myself does not license me to apply 'noncomposite' instead, because it is one instance of my inability to apply to myself—to my 'thinking subject'—any concepts at all. In the following, 'categories' can for present purposes be read as 'concepts':

S does, indeed, occupy itself with an object of experience, but only in that aspect in which it ceases to be an object of experience (B 427).

We can thus say of the thinking 'I' (the soul). . . that it does not know itself through the categories, but knows the categories, and through them all objects, . . . through itself (A 401–2).

The subject of the categories cannot by thinking the categories acquire a concept of itself as an object of the categories. For in order to think them, its pure self-consciousness, which is what was to be explained, must itself be presupposed (B 422).

There is a build-up there: from what can be experienced to

what can be known to what can be thought.

What makes this worth arguing, Kant thinks, is that someone who accepts S for the right reasons might nevertheless mishandle it: one is tempted to 'conclude from the transcendental concept of the subject, which contains nothing manifold, the absolute unity of this subject itself', despite the fact that we 'possess no concept whatsoever' of it (A 340). This temptation generates 'rational psychology'—the invalid derivation of substantive conclusions about the soul from true premises about 'the condition of our knowledge'. In the following passage 'the category of substance' could as well read 'the concept of noncompositeness':

Rational psychology owes its origin simply to misunderstanding. The unity of consciousness. . . is here mistaken for an intuition of the subject as object, and the category of substance is then applied to it. But this unity is only unity in thought, by which alone no object is given (B 421–422).

2

Are there empirical propositions that one might infer from S? I doubt it; but S does have consequences that one might think are empirical.

Consider a piece of rational psychology, closely related to the second Paralogism, by Kant's principal target. Late in his sixth Meditation, Descartes says:

When I consider the mind, that is to say, my self in so far only as I am a thinking thing, I can distinguish in myself no parts; I apprehend myself to be a thing single and entire. Although the whole mind may seem to be united to the whole body, yet if a foot, an arm,

or any other part of the body is cut off, I know that my mind is not thereby diminished.

Descartes's move from 'my mind' to 'the mind' is not (as I alleged in a Note¹ in 1965) a simple non sequitur: it concerns that basic first-person orientation which I have yet to discuss. Another point-missing objection: 'Your mind may have parts that you have failed to distinguish.' Descartes, I think, would reply that nothing could count as discovering that one's mind is composite, because such a discovery would have to take the form '*I* observe that these elements are related thus and so.' This answer is implicit in his reply to a different possible

¹ 'A Note on Descartes and Spinoza', *Philosophical Review* 74 (1965), p. 380.

objection:

Nor can [the mind's] faculties of willing, sensing, understanding, etc. be spoken of as being its parts; for it is one and the same mind which wills, which senses, which understands.

Descartes clearly has in mind something like the Kantian basis for S, and apparently credits S with implications of a kind which Kant rightly denies to it: 'If a foot is cut off. . . '.

Does Descartes think it impossible that some loss of skill, memory, etc. should accompany every physical amputation? That would be an uninteresting mistake; but Descartes probably does not make it, and certainly it is not implied, or even apparently implied, by the Kantian basis for S. What Descartes is declaring impossible, I think, is that some part of his mind should continue to exist in association with the amputated limb while the rest remained associated with the rest of his body. Anyway, I shall use the name 'Descartes' for a possible philosopher who takes that view and mishandles it in ways that make him a suitable foil for Kant. I shall invent my Descartes piecemeal, along the way. Initially, we know only that he adduces S as proving that a mind cannot divide into two.

Two *what?* If we must say 'two half-minds', I am lost. Fortunately, though, we do not need 'half-minds' in order to deal with the main issue or at least with a centrally relevant one. If Descartes denies only that a mind could split into two whole minds, we can still get a Kantian purchase on his position.

Consider the *Smith phenomenon*. Smith's body is halved; each half becomes a complete human body; each of these qualifies fully as the body of a person (as being associated with a mind); there is overwhelming evidence that two dis-

tinct persons (or minds) are involved; and in respect of each of these—call them Smith₁ and Smith₂—there is maximum evidence that he is mentally as well as physically continuous with Smith in the ways that would ordinarily count toward his being (identical with) Smith.

That last clause cannot be contracted to '... maximum evidence that he is Smith'; for Smith₁ and Smith₂, being distinct from each other, cannot both be identical with Smith. We may think of them both as claiming to be Smith, and as completing the form 'Before I lost half my body, I...' with fragments of Smith's biography; but, since identity is transitive, we presumably couldn't endorse all these claims just as they stand. Logical issues arise here,¹ but I needn't discuss them, for my use of the Smith phenomenon is not threatened by any deep problems in identity-logic. We are concerned solely with whether Smith's mind has divided to form the minds of Smith₁ and Smith₂, and that emphasis on division also puts identity-logic in its place: the fission of a mind, if it could happen, would involve the concept of identity in the same way (whatever that is) as the fission of an amoeba.

The Kantian basis for S is obviously consistent with a human body's being divided and its regenerated halves' presenting onlookers with certain patterns of linguistic and other behavior. If Descartes is to be interesting he must allow that the phenomenon might occur while denying that it could be evidence that a mind had divided: 'No-one could know that his mind had divided; therefore nothing is properly describable as the fission of a mind.' Postponing discussion of this inference, consider its premise. Is it true that I cannot intelligibly suppose that my mind might divide into two?

¹ See A. N. Prior, 'Opposite Number', *Review of Metaphysics* 11 (1957): pp. 196–201.

Well, I can make sense of this:- My body is divided into two, leaving me with a half-body which grows into a complete one; the other half develops similarly, and qualifies as the body of a person who claims continuity with the pre-fission JFB; yet I remain confident throughout of my continuity with the pre-fission JFB. In describing a Smith phenomenon with myself as subject, it seems, I must identify myself with one of the post-fission people, and speak of my knowing that there is *someone else* who... etc.

Descartes will say: 'Just so! All you can describe is a division of your body accompanied, perhaps, by the creation of a second mind which resembles yours and is associated with the half of your body that you—*you*—lose.' This is not the abstract, symmetrical, logical point, that JFB₁ and JFB₂ cannot both be identical with JFB. Descartes's point is rather that my description of a Smith phenomenon with myself as subject must be asymmetrical; whence he concludes that a Smith phenomenon could not occur without a relevant, and indeed crucial, asymmetry. He does not say: 'Smith₁ and Smith₂ may both be continuous with Smith; but we may not identify both—and probably oughtn't to identify either—of them with him.' He says, rather: 'Only one can be continuous with Smith, and that one is identical with him.'

This, though not absurd, is wrong. To cash it with a contentful asymmetry we must divorce Descartes's position from the only premises that support it.

We might cash it as follows. If Smith₁ and Smith₂ were observed carefully enough, one would have to find evidence that one of them was an impostor, betrayed by a shifty look; or that one of them was honestly deceived, as shown by curious memory gaps. But Descartes's premises obviously don't imply this. To say that there must be such observable asymmetries is just to say that a strict case of the physical-behavioral Smith phenomenon could not occur.

Descartes's asymmetry claim, then, must say something about the form in which one could experience the Smith phenomenon as subject. But what content of this sort could it have? Descartes may say: 'I have told you. Necessarily, if you undergo a Smith phenomenon only one of the two post-fission minds will be yours—and you will know which it is.' To this I reply: *who* will know? Suppose that Descartes puts his view to Smith just before the phenomenon occurs, and that afterwards both Smith₁ and Smith₂ have the thought: 'Descartes was right! It is quite clear that I am Smith.' Wouldn't their both thinking him right imply that he was wrong?

Descartes may protest, 'They couldn't *both* have that thought'; but he has no arguments to support him. He is certainly not supported by the fact that in describing what a Smith phenomenon would be like 'on the inside' I cannot stay 'on the inside' of both the post-fission people, i.e. by the trivial fact that I cannot intelligibly suppose myself to be two people at once.

Kant is right and Descartes wrong; but I don't deny that the case is queer. Apart from its wild improbability, and from the demands it makes for conceptual modifications, the Smith phenomenon puts us under pressure at a deeper level still. Suppose that Smith knows in advance that plans have been made for Smith₁ and Smith₂ to have lives of, respectively, happiness and misery; and then consider what his attitude can be to this fact. Equipoised calm, because the two prospects cancel out? That doesn't make sense. Elation because he might become Smith₁, mixed with depression because he might become Smith₂? That concedes everything to Descartes. Indifference, because he is due for extinction and won't be involved at all? Perhaps—but can one regard that as obviously, straightforwardly, the attitude that best fits the facts?

3

I have pitted Kant against Descartes on an agreed battleground, namely, the first-person case, with a shared assumption that this is the right place to start. The assumption might be challenged: 'The Kantian basis for S shows that my view of myself must be blinkered; but it doesn't follow that my thinking about myself must be cramped, for I can always listen to what others say about myself and modestly accept it. We might conclude, by majority vote, that every mind is composite.' Kant would find this too glibly dismissive.

The argument might go like this. **Kant:** 'How can you envisage accepting that your mind is composite if you cannot view it as composite?' **Objector:** 'I understand "composite" and "mind" as they occur in a common language, and I understand how someone else's mind can justifiably be called "composite". My mind is just one among others, and I could know that it too is composite—my liberating membership of a community of minds enables me to enlarge my view of myself through others' views of me and my views of others.' **Kant:** 'What you call liberation is really abdication from your intellectual autonomy. You are not giving due weight to the fact that everything you are entitled to say—including the claim that yours is not the only mind—rests ultimately on facts about your mind. You are not entitled to say that your mind is *just* one among others.'

I give Kant that speech on the strength of an important and (I submit) valuable feature of his thought, namely his adoption of what I shall call *methodological solipsism*.¹ The rationale of methodological solipsism is a sequence of near-tautologies. **(1)** Any theoretical problem presents itself to me as a problem only if it presents itself as *my* problem. **(2)** To

solve a theoretical problem of mine, *I* must decide what it is correct or justifiable for *me* to believe about something or other. **(3)** In justifying my beliefs I can appeal only to data that *I* have, and what appeal I make depends wholly on *my* intellectual handling of those data. All these would look sterile but for the emphasis on certain pronouns; but **(1)–(3)**, selectively stressed, do suggest a philosophical program. It is an unfashionable program, I know; but are the above trivialities robbed of their truth by their italics?

Methodological solipsism consists in approaching philosophical problems in ways which are recommended by **(1)** to **(3)** above. It generates certain techniques, prominent among which is an insistence on the first-person test in the philosophy of mind, i.e. upon testing 'There could be a sentient creature which. . .' by asking 'Could I know myself to be such a creature?' Kant's respect for this test explains his initial sympathy with the second Paralogism. The connection can be seen here:

It is obvious that, if I wish to represent to myself a thinking being, I must put myself in his place, and thus substitute, as it were, my own subject for the object I am seeking to consider. . . , and that we demand the absolute unity of the subject of a thought only because otherwise we could not say 'I think' (A 353–354).

A little earlier, he gives a similar underpinning to the Paralogisms as a whole:

I am entitled to say that that which thinks must, in all cases, be constituted as the voice of self-consciousness declares it to be constituted in my

¹ [Note added in 2012.] I apologise for this nine-syllable label. In 1967 it felt all right to abbreviate it to 'MS', but multiple sclerosis should now be accorded exclusive rights to that acronym.

own self. The reason is this: we must assign to things, necessarily and a priori, all the properties which constitute the conditions under which alone we think them. Now I cannot have any representation whatsoever of a thinking being, through any other experience, but only through self-consciousness. Objects of this kind are, therefore, nothing more than the transference of this consciousness of mine to other things, which in this way alone can be represented as thinking beings (A 346–347).

Kant goes too far. It is one thing to value the first-person test, and quite another to disqualify any statement that fails it. We sometimes credit creatures with conscious states that they cannot accompany by an ‘I think’; are such statements shown to be wrong just because, spectacularly, they fail the first-person test? (It is arguable that Kant demands self-consciousness not for all conscious creatures but only for those which make judgments. I think he is committed to the stronger view; but the weaker one raises analogous problems, and my ensuing remarks apply, *mutatis mutandis*, to it as well.)

Kant might reply: ‘No—they are not shown to be wrong, but are shown to be only ways of codifying facts about behavior. What the insight underlying methodological solipsism shows is that “mind”, “conscious”, etc. don’t apply univocally to creatures without self-consciousness and to humans. To say that something is conscious but not self-conscious is to say something totally different from, wholly dependent on, and two levels less basic than, anything one can say about one’s own mind.’ This reply would be rash. To show why, I must take a small detour.

Sometimes in philosophy one wishes to make an essentially first-person point, and to distinguish it from autobiography. Thus one might say ‘My mind is noncomposite, and aren’t you also prepared to say “My mind is noncomposite”?’, trying to *generalize* about the *first* person. This procedure is vaguely described by Kant’s expressions ‘I put myself in his place’ and ‘the transference of this consciousness of mine to other things.’ (These expressions do not *explain* the procedure. To do that, one would have to tackle the neglected problem of elucidating the language in which—extraordinarily—*we discuss with one another* the philosophical problem about other minds.) But the notion of ‘putting myself in his place’ does not describe, even vaguely, the making of specific, contingent, test-passing judgments about the minds of others. These are the topic of my detour. The first-person test allows that someone else could have a headache, but what does Kant say about my judging that someone else *does* have a headache?

He ought to see a problem here. His methodological solipsism, and such statements as that ‘Thinking beings, as *such*, can never be found by us among outer appearances’ (A 357), forbid him to take for granted the ways in which we do—or even the fact *that* we do—apply mental predicates to others. Yet Kant is wholly inattentive to this as to all aspects of the notion of an embodied mind.¹ Strawson notes one significant sentence: ‘[The soul’s] permanence during life is, of course evident, since the thinking being (as man) is itself likewise an object of the outer senses’ (B 415). But there is no evidence that Kant has duly weighed the vital phrase ‘the thinking being (as man)’; and that particular sentence is embedded in, and notably contributes to, a muddle. In one place, Kant does remark that *under certain circumstances*

¹ See my *Kant’s Analytic* (New York: Cambridge, 1966) pp. 56, 143, 222.

‘the thesis that only souls. . . think, would have to be given up; and we should have to fall back on the common expression that men think’ (A359–60). But since the circumstances are not actual or even possible ones, the remark is irrelevant to any third-person judgments that we do or could make. Other apparently relevant sentences occur in A 362–364, but to assess these I should need to understand the surrounding discussion of the third Paralogism.

Sometimes, indeed, Kant is notably evasive:

‘I, as thinking, am an object of inner sense, and am called ‘soul’. That which is an object of the outer senses is called ‘body’ (A 342).

The second sentence’s lack of any personal pronoun looks like a deliberate attempt to avoid any reminder of the problem about how the soul is connected with the body.

Still, what *could* Kant say, consistently with his methodological solipsism, about contingent, test-passing, third-person statements about minds? Almost anything. Methodological solipsism implies a partial program for seeking an analysis of third-person statements, but makes no prejudgment about what must be found. In stressing what *I* can know given *my* data it does not imply, for example, either (a) that I could be self-conscious even if there were no other minds or (b) that I could be self-conscious even if I had no perceptions of an outer world. For what the fact is worth: I am inclined to accept (a), but only because the arguments so far adduced against it seem to me invalid; and I reject (b) on the strength of a thoroughly methodological-solipsist argument against it. Methodological solipsism does perhaps imply that some of my first-person statements are in some way more basic than any of my third-person statements; but that says almost nothing.

The detour is complete. We are confronted by **(1)** first-person judgments, **(2)** test-passing third-person judgments,

and **(3)** test-failing judgments; and we are considering a methodological solipsist who, seeing that none of **(3)** could be expressed as members of **(1)**, avoids dismissing them all as false by saying that they aren’t what they seem. The members of **(3)**—he says—are not really about minds at all but about behavior-patterns, and statements expressing them use mental predicates with meanings other than those they have in statements expressing **(1)** and **(2)**. But this implies an entitlement to separate **(1)** and **(2)** from **(3)**, and that won’t do at all. It isn’t clear that a methodological solipsist can avoid behaviorism about **(3)**, but the reasons for thinking that he can’t are also reasons for thinking that he can’t avoid behaviorism about **(2)** either. If he is to isolate **(3)** by saying that they admit of a purely behavioral analysis, then he must show—not assume—that behaviorism is not the whole story about **(2)**. I don’t deny that this can be shown; but attempts so far have satisfied few but their authors, and Kant in particular doesn’t even make one. (I have encountered a tendency to associate the line separating **(1)** and **(2)** from **(3)** with the line separating creatures that use language from ones that don’t, and to infer that **(3)** is therefore on a different footing from **(2)**. But even if that association held—which it doubly fails to do—it is not obvious that it would save **(2)** from behaviorism: language is linguistic behavior, and the question of what it is to understand what another says, like every epistemological question, can be answered thoroughly and in depth only by a methodological-solipsist approach.)

I conclude that Kant is entitled to his initial sympathy with the second Paralogism: there are sound reasons for exploring what might be called the ‘logic of the first-person case’. But those reasons imply that there are problems about **(2)** test-passing judgments about the minds of others and **(3)** test-failing judgments. Kant fails in his obligation to face up to those problems—in the case of **(2)** by evasion, and in the

case of **(3)** by an implied dismissiveness to which he has not earned the right.

The fourth part of my paper is merely a coda. It seeks

Of all great modern philosophers, Spinoza is probably the least given to methodological solipsism. He should, but does not, ask ‘How, if my philosophy is true, can I know that it is?’. He does not, and cannot, give a remotely satisfactory account of self-consciousness. His system’s denial of any special status to ‘I’ is as damaging as its more notorious dethronement of ‘now’. These charges point to some of the gravest weaknesses in Spinoza’s remarkable edifice. But if his work loses depth by its freedom from methodological solipsism, it also gains scope. In particular, it makes room, as the Cartesian and Kantian contexts do not, for worthwhile questions about **(3)** judgments that fail the first-person test.

In Part II of the *Ethics*, for example, the material between propositions 13 and 14 implies that the notion of an individual body is neither sharp nor absolute: an ordinary human body, for instance, is a complex configuration that we can regard as ‘one body’ only because its parts interact in ways that give it a sufficient degree of unity. (This, together with Spinoza’s psycho-physical parallelism, yields proposition 15: ‘The idea which constitutes the actual being of the human mind is not simple, but compounded of a great number of ideas’, which tramples down the second Paralogism and even Kant’s limited sympathy therewith. But that is by the way.) That suggests that the everyday conceptual opportunism whereby we count bodies might sometimes fail us because some physical phenomena remain obstinately borderline between ‘one body’ and ‘two bodies’.

to widen the historical setting, to amplify some of the foregoing remarks, and—finally—to add a suggestion about the analysis of **(2)** test-passing third-person judgments.

4

And that raises a further question: could it ever be correct to say ‘This organic configuration is precisely borderline between being a single creature with a mind and being a pair of creatures with a mind each’? An affirmative answer clearly fails the first-person test; but the question still seems to me an open and an interesting one, and I am grateful to Spinoza for raising it.

Objection: ‘A borderline case between “an organism” and “a colony of organisms” need not also be borderline between “a mind” and “a colony of minds”’. Spinoza thinks otherwise, but only because of his wholly implausible psycho-physical parallelism. Aren’t you implicitly congratulating him upon that?’ No, I reject Spinoza’s thesis that everything extended has its mental counterpart, related to it as the human mind is to the human body. But that thesis is rooted in two good insights: **(a)** Our non-first-person judgments about what minds there are and what they are like must be closely linked to some of our identifications and descriptions of bodies; and **(b)** On a scale of known kinds of body, from inorganic through to human, we cannot make a precise, absolute, nonarbitrary cut at the point where mentality comes in. These defensible theses are turned into unrestricted psycho-physical parallelism simply by Spinoza’s rationalistic passion for depth, definiteness, and theoretic tidiness, which turns **(a)**’s ‘closely linked to some’ into ‘wholly paralleled by all’, and **(b)**’s ‘we cannot make a precise etc. cut’ into ‘there is no difference of kind between the two ends of the scale’. This

is not a defence, but a representation of Spinoza's position as consisting in some valuable insights which—understandably but not inevitably—he has extravagantly inflated.

This picture at least gives us hope of learning something from the viable core of Spinoza's philosophy of mind. A borderline between 'one' and 'two'—to return to my example—can be sought in possible organic configurations to which mental predicates could be plausibly applied. To be led by Spinoza to the thought of a borderline between 'one mind' and 'two minds', one need not swallow his psycho-physical parallelism whole.

What one does need, though, is freedom from the inhibitions induced by the first-person test.

Spinoza bought his freedom from those inhibitions at too high a price. One's attention to **(3)** test-failing judgments may be paid for not by ignoring **(1)** the first-person case but merely by failing to relate **(3)** fully to it. This is unsatisfactory too, but it improves on Spinoza and may be the best we can do in the meantime. Work done on **(3)** in this spirit will be thoroughly, if provisionally, behavioristic (I use this term throughout in a very broad sense). Pending a satisfactory relating of **(1)** with **(2)** and of both with **(3)**, we shan't know whether behaviorism is fully adequate for **(3)**, nor shall we know whether its adequacy would isolate **(3)** from all our other judgments about minds. But that can be borne with, so long as we know—as we surely do—that a wholly behaviorist approach to **(3)** can yield results which are not seriously wrong in detail.

Provisional behaviorism also lets us proceed with the analysis of mental concepts as used in **(2)**, but here we must tread more delicately.

The case for thinking that behaviorism is the whole story about **(2)** stems from a persuasive and usually helpful view about the relationship between what I can mean and what

data I could have. The most persuasive case against it depends upon a denial that behaviorism will do for the first-person case, together with a denial that mental terms are systematically ambiguous as between **(1)** and **(2)**. In short, the unsolved problem that some of us have takes this form: we cannot reconcile our views about meaning in general with our views about the meanings of **(1)** and **(2)** in particular. I suspect that our views about meaning in general must yield, or broaden; but I have nothing useful to say about that. What I do want to suggest is a way of reducing the area of conflict.

Briefly, the suggestion is that behaviorism may be adequate for all our *distinctions* between mental states. Even if no set of hypotheticals about behavior etc. can entail 'He is in pain', such a set might entail 'If he feels anything, he feels pain' and 'He is not wholly absorbed in a feeling of mild euphoria'. This is not to propose that we partly separate **(2)** from **(1)** by conceding behaviorism for one aspect of the former, but rather to suggest that behaviorism is adequate for *that* aspect of both.

A wholly behaviorist account of **(1)** seems unacceptable. I think there is more to my being in pain than just my behaving and being disposed to behave in certain ways, even if I cannot intelligibly elucidate 'more':

The very fact that we should so much like to say: 'This is the important thing'—while we point privately to the sensation—is enough to show how much we are inclined to say something which gives no information (Wittgenstein, *Philosophical Investigations* § 298).

Perhaps; but it also shows that behaviorism is felt to be inadequate for the first-person case. For those who feel this, the great unsolved problem is, precisely, to 'give information' about the inadequacy or to show that what they now 'like to say' does after all 'give information' that isn't captured

by a behaviorist analysis. (Their position is intellectually disreputable—I used ‘feel’ advisedly—and I venture to confess my unargued sympathy with it only because I think that many others are similarly inclined. Only they could be interested in my proposal.) The desire to say ‘This [pointing privately] is my pain’, then, solves no problem; but it locates a problem, and so it should be distinguished from the desire to say ‘It is because of the nature of this [pointing privately] that I count as being, specifically, in pain’. The first desire need not be accompanied by the second. To give expression to the first alone would be so say: ‘My being in pain is my having a conscious state such that I (am disposed to) behave thus and so.’ The phrase ‘conscious state’ merely marks the point where resistance to behaviorism begins, but if all such resistances could be expressed in that form then something would be gained. For then we should have only the unitary problem of making better-than-behaviorist sense of the notion of ‘conscious state’ as such; and the elucidation of differentiae—of statements attributing this

rather than that conscious state to any creature—could proceed along wholly and non-provisionally behavioristic lines. The second half of ‘conscious state *such that*’ probably need not be construed causally; but if it is to do any work at all in my suggested formula, then the suggestion mustn’t be that any two conscious states differ only in the behavior etc. in which they are manifested. Room must be left for ‘inner’ differences, which can’t be described behavioristically. That, however, leaves the suggestion standing: it may be that such ‘inner’ differences play no part—or only a part depending entirely on their association with behavior etc.—in applications of mental predicates to anyone, even oneself.

Teachers sometimes lead into these matters by saying ‘How do you know that his feelings are not utterly belied by his outward appearance?’, or by saying ‘How do you know that, although he is outwardly responsive, he is not all blank within?’ If my tentative suggestion is right, these are not simply less and more radical introductions to a single problem.