

Henk. Johan Glock Wittgenstein: A
Critical Reader (2001, Blackwell: Oxford)

11

Wittgenstein and 'I'

David Bakhurst

1 Introduction

It is a truism that we bear relations to ourselves unlike the relations in which we stand to others. We are, for example, aware of many of our mental and bodily states in a way in which others cannot be. You and I may both attest that I am in pain, but I *feel* my pain while you observe its manifestation. My pain is part of what is sometimes called my 'inner life', a metaphor that signals the fact that my conscious experiences are not just isolated events, but are united under the distinctively first-personal perspective of an enduring being. This perspective involves self-consciousness: I am aware of myself and my relation to things around me, of the continuity of my existence, and of my own awareness.

Deep philosophical issues arise as soon as we seek to explicate these truths. What is the character of our self-awareness? Is it an awareness of an object or entity that is ourselves (i.e. awareness of a 'self')? What are the criteria for the identity of a self or person through time? Is the special relation the self bears to its states something like possession or ownership, or should it be conceived some other way?

This chapter discusses Wittgenstein's treatment of these and related issues. I shall focus on passages brought together in the section of *The Wittgenstein Reader* entitled 'The First Person'. These selections comprise the well-known discussions in the *Blue Book* on personal identity and the status of the first-person pronoun, further reflections on the latter topic from the *Philosophical Investigations*, and passages about various psychological phenomena (such as sensation, belief, consciousness and introspection) that illuminate first-person psychological ascription in particular and self-consciousness in general.

Wittgenstein and 'I'

225

Although these selections may strike the reader as disparate, disjointed and obscure, they have an important place in Wittgenstein's later philosophy. The material complements the critique of broadly Cartesian conceptions of the mind developed in Wittgenstein's more tightly woven discussions of rule-following and private languages. His reflections on the first person continue to attack the idea that each 'self' inhabits a kind of self-contained world of private thoughts and experiences, together with the Cartesian tenet that the mind's awareness of its states constitutes a privileged form of knowledge, an idea exploited not only by Descartes, but by many subsequent thinkers (including the British empiricists, Russell, and prominent logical positivists). The material is also interesting for the light it casts (or fails to cast) on the alternative conception of thinking beings that begins to emerge in the discussions of rule-following and private languages. There Wittgenstein urges us to abandon the idea of the mind as a private theatre of representations and, it seems, embrace the view that each minded being may manifest its mental powers only in a public world and in interaction with other such beings. It is important to consider whether Wittgenstein's remarks about the self help clarify and develop this alternative vision.

2 Personal Identity

The nature of personal identity seems a promising place to begin a discussion of the first person. As we noted, the first-personal perspective is the perspective of an enduring being. Persons are such beings.¹ It therefore seems appropriate to inquire into what is involved in the continued existence of a person through time. To approach that question we ask: what makes a person at two different times the *same* person?

Philosophers have typically answered the latter question by proposing a criterion of personal identity stating necessary and sufficient conditions for persisting through time as the same person.² One popular view, which has its origins in Locke's *Essay*,³ casts the criterion in terms of psychological continuity. The latter notion needs careful handling. Locke supposed that *B* is the same person now as *A* was at some earlier time if and only if *B* has memories of *A*'s experiences. But this is too strong a requirement: people often lack experiential memories of their infancy and of many later events in their lives. A more reasonable view is that psychological continuity requires only that there be an adequate number of direct links between a person's psychological states at each adjacent stage of her existence. So, *B* now is the same person as *A* was then if and only if *A* and *B* are linked by a sufficient number of overlapping chains of directly connected psychological states.

Against such a view, some philosophers argue that bodily continuity is the criterion of personal identity. *B* is now the same person as *A* was then if and only if *A*'s body – or enough of it to sustain a living person – has continued to exist and is now *B*'s.

Discussions of these criteria, in their various versions, often make two assumptions. The first is that questions of personal identity have determinate answers. When we consider the possible consequences of my being subjected to an operation that might change me in radical ways, we assume that whether it will be I or someone else who wakes up after the operation is a question with a determinate answer. Survival is an all-or-nothing matter. I will either survive the operation or not; there is no third option. The second assumption is that criteria of personal identity can be best defined by conducting thought experiments and by reflecting on unusual or abnormal 'real-life' cases. Locke, for example, argues for his view by invoking the possibility of a person changing his body. He writes:

Should the soul of a Prince, carrying with it the consciousness of the Prince's past Life, enter and inform the Body of a Candler as soon as deserted by his own Soul, every one sees, he would be the same Person with the Prince.⁴

Cases of multiple personality disorder are also adduced to show that, since we countenance the possibility of two persons inhabiting the same body, the criterion of personal identity must be psychological continuity.⁵ In reply, bodily continuity theorists invoke cases, real and imagined, in which we suppose a person to survive radical psychological disruption. Total amnesia, we are reminded, is not the death of one person and the birth of another. And if you were to be captured by a mad scientist who proposed to 're-format' your brain and replace your present mental states with different ones, you would surely see this operation as something that would be about to happen to *you* and from which *you* would emerge changed psychologically beyond all recognition. If so, psychological continuity cannot be essential to personal identity.

That such problem cases generate apparently compelling arguments for both criteria suggests that we have an antinomy. In fact, the matter is more puzzling, for there is a powerful argument against both psychological- and bodily-continuity theories. Consider the possibility of the division of a person into two.⁶ There are two cerebral hemispheres in the brain. There is some division of labour between them (e.g. the left hemisphere controls speech production, visual information from the right eye goes to the left hemisphere (and the converse), the left hemisphere controls the right hand (and the converse), but in the event of damage to one hemisphere, the other is sometimes able to learn

its functions. It seems that either hemisphere could sustain a living person. Moreover, we know from real cases in which the corpus callosum is surgically severed, dividing the two hemispheres, that patients can be made to exhibit behaviour that suggests the presence of, as R. W. Sperry put it, 'two independent streams of conscious awareness, one in each hemisphere'.⁷ So let us imagine a case where the brain of a person, *P*, is entirely divided and each half transplanted into a different body. Immediately after this operation we have two beings, *Q* and *R* who, by both psychological and bodily criteria, have an equal claim to be *P*. Both *Q* and *R* are psychologically continuous with *P*, and in both cases there is also continuity of enough of *P*'s body to sustain a living person (i.e. half *P*'s brain). But *Q* and *R* cannot both be *P*.

How would Wittgenstein have responded to such a treatment of personal identity? His remarks in the *Blue Book* suggest he would have had little patience with it (BB 61–2/WR 191–2). He begins by reminding us of the criteria we actually use to identify and re-identify people: their bodily appearance, characteristic habits and behaviour, the sound of their voice, and so on. The suggestion here is that it is a mistake to think there is a single criterion of personal identity. Rather, the term 'person' has a 'composite' use, governed by many overlapping criteria. Moreover, that we can deploy such criteria – that we have the concept of personhood we do – depends on certain contingent facts, such as the fact that a person's appearance and characteristic ways of behaving tend to change only very slowly. Were circumstances different, Wittgenstein attests, our practices of naming and identification would likely be different too. If human beings all looked alike but certain psychological characteristics and modes of behaviour migrated among their identical bodies, we would no more be inclined to associate names with bodies than we are to name our dining-room chairs (BB 62/WR 191). Our very idea of discrete persons who persist through time is thus premised upon certain contingent facts.

It is thus a mistake, Wittgenstein argues, to think that deep metaphysical truths underlie our concept of a person. If we were 'shape-shifters', alternating between two different physical forms with accompanying and different psychological states, we would likely think of each shifter as two people. But nothing forces us into this conclusion. The terms 'person' and 'personality' are made for use in ordinary, everyday contexts. Nothing dictates how their use should be extended in the extraordinary ones. It is not that these terms have, as Wittgenstein puts it, 'only one legitimate heir' whose authenticity can be established by philosophical analysis (BB 62/WR 192). How we extend their use is a matter of judgement.

Wittgenstein thus rejects the presuppositions of so much philosophical writing about identity. He denies that there is a single criterion of personal identity to be cast in terms of *either* psychological *or* bodily factors. There are

no strict necessary and sufficient conditions of personhood. He explicitly denies that we can learn substantive truths about identity by reflecting on puzzle cases. He denies that all questions of identity have a determinate answer, and he would no doubt countenance the possibility that we might revise our view of survival as an all-or-nothing matter.

It is worth comparing Wittgenstein's position to other views that attack traditional assumptions of the debate. Consider the position Derek Parfit ingeniously develops in *Reasons and Persons*.⁸ Unlike Wittgenstein, Parfit is confident that we can establish and refine criteria of personal identity by philosophical analysis and he makes extensive use of puzzle cases. What Parfit takes his analysis to show, however, is that it is not the case that all questions of personal identity have determinate answers. (He argues like this: imagine that *X* is subjected to a series of gradual changes to both bodily and psychological characteristics which result in an individual, *Y*, who cannot be seen as the same person as *X*; there is no firm line within this spectrum of changes at which *X* ceases to be and *Y* comes into existence.)⁹ Parfit then argues that questions of identity are not what really matter. The important relation between us as we are now, and as we will be, is psychological continuity, which we must distinguish from identity because of the possibility of an individual dividing to form two selves both psychologically continuous with her undivided self. Parfit proceeds to suggest that, even in actual, familiar cases, it is sometimes appropriate to think in terms, not of single enduring selves, but of relations between successive selves. Thus I might see my relation between myself as I am now and myself as I was as a ten-year-old as a relation between my 'present self' and one of my 'earlier selves'. Whether this way of talking is appropriate depends simply on the extent to which there is significant discontinuity between myself as I am and as I was.¹⁰

Parfit is methodologically less radical than Wittgenstein, but the resulting position is more radical. For Parfit, philosophical analysis reveals substantive truths about identity that compel us to revise our view of the kind of beings that we are, and, moreover, to do so in a way that has significant ethical implications. Parfit is explicitly engaged in what Strawson calls 'revisionary metaphysics'; he aspires to challenge and transform our existing concepts.¹¹ Wittgenstein, in contrast, denies that there is anything wrong with our everyday concepts. We get paradoxical results when we consider unusual cases, real or imaginary, not because our concepts are faulty but because they are made to work in the domain of the ordinary. We should think of concepts as tools fit for specific purposes (BB 67–8/WR 193–4), and not suppose them inadequate because they fail us when we try to use them for bizarre jobs. If the bizarre were to become the norm, we would no doubt have to refashion our tools. But that is no reason to refashion them now.

It is also worth contrasting Wittgenstein's persons developed by Daniel Dennett in *Cartesianism*. Dennett denies that there is an abandon the idea that, when it comes to self to a Customer,¹² and admit that

selves are not independently-existing soul processes that create us, and, like other such status.¹³

Selves, or persons, are posits or 'fictions', 'gravity' of an object. Indeed, Dennett put describing the self as a 'centre of narrative' to deploy in weaving a set of experiences and single life.

There are certainly elements in Wittgenstein's constructivism.¹⁴ It is true that, for him, the concept of a person is more than facts about how. Indeed, Wittgenstein's claim that, were the we use the term 'person' to change, we would use might suggest a constructivism more (62/WR 192). This impression is misleading linguistic practice, nor 'reality', could dictate should be revised, certain ways of speaking natural to us in the novel situation. Though pher to be an unconstrained choice, the practical realities of our life may force up than another.

The real difference between Dennett and does not treat the notion of a person as a concepts. Our physical concepts are equally are not 'abstractions' or 'fictions' constructs. Hence we do not need what Dennett of the self as a centre of narrative gravity) the create character of our selves while enabling natural order. In contrast to both Dennett attempt to use philosophy to revise our view about the nature of persons. On the contrary confusions that surround the concept of a person a proper appreciation of the fact that there is a problematically problematic about persons at all.

3 The Bare Self

Wittgenstein, Parfit and Dennett do share something important: the idea that questions of personal identity may be indeterminate. This view is fashionable in the contemporary literature, yet some nagging intuitions speak against it.

Take the case where a scientist proposes to fiddle with my brain so that I acquire psychological states discontinuous with those I presently have. As my description of the case betrays, I conceive of this as something that will happen to me. I can imagine myself surviving this radical psychological transformation, even though after the operation I will have no memories of myself as I was. I can also imagine a scientist transplanting my brain into a new, perhaps even artificial, body, and then gradually replacing my brain's 'circuits' with artificial 'neurons'. Indeed, it seems I can imagine the scientist proposing to alter me radically in both psychological and physical respects, so that after the operation I would have no recall of my past psychological states and would think, behave and look like an entirely different person.

Admittedly, when we consider such experiments from a third-person perspective, we are inclined to describe them differently. From that perspective, you will probably say that in each of these experiments I cease to exist and that the question of when exactly I do so cannot be answered. Yet why privilege the third-person perspective? It is, after all, the first person that we are trying to understand. And, as Bernard Williams reminds us, from the first-person perspective, what I am primarily interested in when I consider my prospects of surviving such an operation is whether I will continue to exist as a subject of experience.¹⁵ I ask myself whether the post-operative pain the survivor will feel will be felt *by me*. Will it hurt me, or will there be no me to hurt? When we put this question to ourselves, we feel it must have a determinate answer.¹⁶

It appears that a crucial part of our existence as persons involves existence as subjects of consciousness, as points of self-conscious awareness, and it seems intelligible that we should endure as such despite radical changes to our psychological and bodily states. Parfit dismisses such a view as Cartesian: to conceive of such 'bare selves' is to admit the existence of Cartesian egos. Williams, in contrast, takes the view to support bodily-continuity theory. If personal identity involves continuity of existence as a subject of consciousness, then it is reasonable to suppose that the latter continuity depends, albeit contingently, on brain functioning (though the possibility of division remains problematic). But to portray the bare self as physically instantiated in the brain is, Parfit and Dennett would counter, no less Cartesian (Dennett speaks of 'Cartesian

materialism'),¹⁷ for the self is portrayed as a set of capacities logically independent from the mental states those capacities realize, and of the physical states in which those capacities and states are embodied.

Cartesian or otherwise, the intuition that the core of our being is a mode of self-conscious awareness that might persist through radical psychological or physical disruption is a powerful one. Wittgenstein's reflections on the grammar of the first-person pronoun constitute an attack on this idea of the self (and indeed, on any view which represents the self as an object) rather different from those we find in Parfit and Dennett. On the bare-self model, the self, as subject, is seen as an entity, as the *possessor* of psychological and bodily states. It is thus tempting to analyse first-person psychological ascription ('I feel pain', 'I am seeing a red patch', 'I believe it is raining') in the following way: 'I' refers to the self and the psychological ascription describes or reports mental states possessed by that self. Wittgenstein attacks both sides of this picture. 'I', he argues, is not a referring expression, first-person psychological ascriptions are not reports or descriptions, and the relation between a person and his mental states cannot be conceived as one of possession or ownership.

4 Wittgenstein on the Reference of 'I'

There is controversy in philosophy about what is to be included in the class of genuine referring expressions, or 'singular terms'. A relatively uncontroversial definition is that an expression is a referring expression if its role is to indicate some particular object or entity. Thus the class certainly includes proper names ('Ludwig Wittgenstein', 'London'). Demonstrative expressions ('that dog', 'this cat') are also often taken to be referring expressions, as are definite descriptions ('the rudest man in Europe'), though there is controversy about the latter.

Let us focus on proper names. Our use of a name, for example 'James Jones', is governed by criteria of identification. We identify James by his properties (by the very kind of property that Wittgenstein points to in his discussion of personal identity). When someone asks, 'Someone is making a racket in the next room, is it James Jones?', we use such criteria to determine the answer. In such cases, of course, we can misidentify who is making the noise. We can take someone else for James. It is also possible to use a name, or what one takes to be a name, and simply fail to refer to anyone.

In contrast, uses of 'I' seem immune to error through misidentification or failure of reference. When I think 'I am in pain', I cannot be wrong about whom I have in mind, nor can there fail to be such a person. Thus, if 'I' were a referring expression, it would be guaranteed correctly to refer

on each occasion of its use. This, for Wittgenstein, is a reason to think that it is not a referring expression at all.¹⁸

Is 'I' immune to error through misidentification in *all* its uses? In the *Blue Book* Wittgenstein argues that immunity applies only in cases of first-person psychological ascription, where, as he puts it, 'I' is used 'as subject' (BB 66–70/WR 192–6). There is no immunity, he claims, where 'I' is used 'as object', that is, when I ascribe bodily states to myself. Someone who claims, say in the aftermath of a car accident, 'I have broken my arm', can be wrong because she takes the broken arm of a passenger beside her to be her own – an error due to misidentification.

Wittgenstein's distinction is confused and he later abandoned it. Even when 'I' is used 'as object' there is no chance of my misidentifying myself in the sense of taking myself for someone else. In a case where you have ~~broken~~ your arm and I mistakenly judge that it is I whose arm is broken, I do so because I mistake your injured arm for my own, not because I mistake myself for you. I am not confused about whom I suppose to have broken my arm.¹⁹

Consider the case where we wondered whether James Jones was making a racket next door. There are two sorts of misidentification that can confound us here. First, knowing perfectly well who James is, we conclude he is next door, when in fact it is someone else. Second, we can be wrong about who James Jones is; that is, about what James Jones signifies. Uses of the first-person pronoun are immune to error of the second sort, whether they involve the ascription of psychological or bodily states.²⁰ I cannot be confused about what object 'I' signifies when I use it, even if I am confused about 'who I am' because I am amnesiac or deluded. The amnesiac's confusion in no way interferes with her ability to use the term 'I'.

The example of the amnesiac shows, of course, that the use of 'I' is not mediated by criteria of identification. I need know (almost) nothing about myself in order to refer to myself in the first person. I may be unable to recognize myself in the mirror, but things cannot be so bad that I ask 'Who am I?' and then misidentify the 'I' whose properties I want to know.

All this suggests to Wittgenstein that 'I' does not serve to identify a person at all. When I think and speak to myself, I do not deploy the term 'I' to pick myself out from others. It might be argued that the term does have an identifying function in communication (e.g. "'Who wants ice cream?', 'I do!'). But Wittgenstein denies even this. The term does not identify me, first, because anyone can use it, and second, because someone can fully understand my use of 'I' even though they know none of my identifying features. The term does not point to anything, any more than a groan points to the groaner. Its use, in such public contexts, serves rather to draw attention to ourselves.

We might grant, however, that the first-person pronoun is not, as Julie Andrews once sang, a name I call myself, and yet maintain that it refers. 'I' is, after all, an indexical; the use of other indexicals – 'this', 'that', 'here' and 'now' – is not governed by criteria of identification, but they are surely referring expressions. Why not treat 'I' the same way?

Although there are no general criteria of identification associated with 'this' and 'that', their use requires that the speaker intend to single out some particular object(s), and in order to do this, she must have some grasp, however momentary and inarticulate, of what distinguishes the object(s) in question (consider, for example, a speaker using 'this' to refer to a coloured patch; for the reference to be successful, she must be able to distinguish the patch from its background). And though I need not know where I am, and when it is, to assert correctly, 'I am here now', uses of 'here' and 'now' still require the speaker to define their range. Whether 'now' refers to a moment or an era, and 'there' to a point or a galaxy, must be established by the speaker. What makes Wittgenstein suspicious about 'I' is that neither speakers' intentions, nor contextual considerations, are ever relevant to its successful use.

But why not conclude that 'I' is a super-referring expression? If attempting to refer to something is like firing an arrow at a target, we can think of 'I' as a magic arrow that cannot miss. Wittgenstein rebuts this suggestion. Using 'I' is more like drawing a target around an arrow already stuck in the wall. Where a term is used to refer, some work must be done to pick out a particular object. This is not so with the use of 'I', and thus, Wittgenstein concludes, we should not think of it as a referring expression at all.

5 Wittgenstein Criticized: Persons Again

I do not find Wittgenstein's position wholly plausible. There is, after all, a prima facie case for treating the first-person pronoun as a referring expression. To deny that 'I' refers is, as Wittgenstein himself admits, paradoxical (MS 116, 215). 'I' is a logical subject; it is a noun-phrase, and one which can be replaced *substantive* by expressions (names) that clearly do refer to the speaker.

The denial is made to look plausible by Wittgenstein's use-centred view of language. Wittgenstein sees referring as an *act*, as something we attempt for the purpose of 'identifying' or 'singling out' objects, and at which we may either fail or succeed. It thus seems that no purpose could be served by the subject's referring to herself, especially in thought. She does not, after all, need to remind herself of her presence, or to 'pick herself out'.

One might counter, however, that the primary function of referring expressions is to establish the object about whom we are thinking or speaking. Why

deny 'I' such a role? The fact that a person, when thinking of herself, does not need a device for singling herself out in thought is not grounds to conclude that 'I' does not refer for two reasons. First, no thinker is alone in the world. The first-person pronoun is part of a public language and in communicative contexts the function of 'I' is (at least in some sense) to indicate the individual who is speaking. Second, the first-person pronoun is deployed in the expression of thoughts, feelings and desires, and if I am to express myself I must, as it were, establish a first-personal perspective. There is an indexicality to this perspective that is ineliminable.²¹ It is *mine*. Of course, I recognize an equivalence between the statement, 'I am giving a lecture', said by me, and the statement, 'David Bakhurst is giving a lecture', said by anyone. But I recognize the equivalence only because I accept as true a further indexical thought – that I am David Bakhurst. If the first-person pronoun is a *critical* ingredient in articulating such first-personal thoughts, then its use has a purpose, and one not somehow undermined by its immunity to referential error. Why not suppose that it achieves this purpose by referring to the thinker whose thoughts they are?²²

Wittgenstein may have felt that an additional reason to resist this conclusion is that there is no suitable object to serve as the reference of 'I'. Elizabeth Anscombe, in her classic article 'The First Person', which seeks to defend Wittgenstein, argues that since 'I' can be used by someone in a condition of extreme sensory deprivation, i.e. unaware of her body, it cannot refer to the body. Loss of body is not loss of self. Thus, she concludes, if 'I' refers, it refers to something bodiless.²³ If we treat 'I' as a referring expression, we find ourselves committed to the existence of Cartesian egos.

In the *Blue Book* Wittgenstein appears to make a similar point:

We feel then that in the cases in which 'I' is used as subject, we don't use it because we recognize a particular person by his bodily characteristics; and this creates the illusion that we use this word to refer to something bodiless, which, however, has its seat in our body. In fact *this* seems to be the real ego, the one of which it can be said, 'Cogito, ergo sum'. (BB 69/WR 195–6)

This is a perceptive passage. But to deny that 'I' refers to something bodiless is not to deny that 'I' refers at all. That we can be aware of ourselves in a way that is not an awareness of any of our bodily features does not show that we are essentially featureless things. Indeed, it does not even show that we are not essentially embodied beings, unless it is impossible that an essentially embodied being should have a mode of self-awareness, or self-presentation, unmediated by criteria of identification.

It is easy to suspect that a residual dualism infects both Wittgenstein's and Anscombe's treatment of the issue. It seems that *either* 'I' refers to the body (which it obviously does not, since 'I am a Spurs supporter' does not mean 'My body is a Spurs supporter') or to something immaterial (which is philosophically unacceptable).²⁴ So we have to conclude that 'I' does not refer at all. But there is an acceptable alternative. We can hold (contra Wittgenstein at LPE/PO 228) that 'I' refers to a *person*, and we can follow Strawson and treat persons as living creatures with certain distinctive mental powers.²⁵ Persons are neither bodies nor bodiless selves, nor combinations of the two, but embodied beings with distinctive psychological capacities. We can now say that 'I' refers to the person who utters it in speech or thought,²⁶ and we can treat this as what Wittgenstein calls 'a grammatical truth', as one of the deep conventions of our language. Of course, it always looks as if grammatical truths have a metaphysical warrant, so it is hardly surprising that discussions of the first person are a swamp of philosophical misconceptions. But things are not so desperate that we have to deny that 'I' refers.

Let us briefly consider three questions that immediately present themselves. First, if we think of persons in this Strawsonian manner, what becomes of the criteria of personal identity? Let us say that to endure as a person over time requires the continued existence of a body necessary to sustain a living being who possesses psychological capacities able, in turn, to sustain a mental life that exhibits appropriate psychological continuity. In the spirit of the *Blue Book*, let us think of this as specifying a norm drawn from our everyday practice in normal circumstances, rather than as legislating across logical space. The norm does not fix the extension of the concept to abnormal circumstances, nor does it guarantee that what we now count as normal will continue to be so.

Second, if we cannot dismiss the idea that 'I' refers to an entity, how are we to undermine the intuitions which drove us to entertain the idea of the bare self? We can now see that this idea gets its life from the illusion Wittgenstein diagnoses in the quotation above. Because our awareness of self is not mediated by criteria of identification, it seems intelligible that we may endure some operation which radically transforms our psychological and bodily states. It seems that our bare self-awareness may persist come what may. But that appearance does not determine what has to be true for such self-awareness to be possible and for it to endure. It is open to us to conclude (though not on grounds of introspection), that only an embodied being could enjoy such self-awareness, and hence the illusion of being a bare self.

Third, are we now forced to embrace a view of persons similar to Parfit's? I do not think so. Parfit advances the 'reductionist' claim that a person's identity over time just consists in the holding of certain facts which can be

described in an entirely impersonal way, without presupposing the identity, or even the existence, of that person.²⁷ Yet in so far as such facts are facts about mental states (e.g. facts about psychological connectedness and continuity), it is unclear whether such an impersonal description is possible, since, as Strawson argues, mental states 'owe their identity as particulars to the identity of the person whose states or experiences they are'.²⁸ Moreover, although personhood requires bodily continuity sustaining the right kind of psychological continuity, we conceive of a person as something more than a collection of gradually changing states.

Our concept of a person is a concept of the subject of a life, a locus of capacities and possibilities, the focus of a distinctive point of view. One wants to say (with Hegel in mind) that if there is a sense in which I am identical with my evolving states, there is also a sense in which I am not identical with myself. We cannot appreciate the flux of our being, the magnitude of our becoming, without a robust sense of ourselves as *that-which-becomes*.²⁹

6 Avowals: Giving Voice to the Mind

Wittgenstein was, of course, by no means the first philosopher to attack the idea of the self as an entity. Hume famously argued that

when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other of heat or cold, light or shade, love or hatred, pain or pleasure. I can never catch *myself* at any time without a perception, and never can observe anything but the perception.³⁰

In the *Tractatus* Wittgenstein makes a similar remark:

The thinking, presenting subject; there is no such thing.

If I wrote a book, 'The world as I found it', I should have therein to report my body and say which members obey my will and which do not, etc. This then would be a method of isolating the subject or rather of showing that in an important sense there is no subject: that is to say, of it alone in this book mention could *not* be made. (5.631/WR 25)

The *Tractatus* proceeds, in Kantian style, to adopt a transcendental view of the metaphysical subject as a limit of the world rather than a presence within it. But in the immediately post-*Tractarian* period, Wittgenstein advanced a position with even greater similarity to Hume's 'no-ownership' conception of the mental (see M 101).³¹ Embracing a form of methodological solipsism,

Wittgenstein argued, following Lichtenberg, that 'I' is redundant. For the solipsist, there is pain' serves as well as 'I am in pain'. The first-person perspective is eliminable.

In the later philosophy, Wittgenstein famously abandoned the solipsistic perspective of his earlier work. As we have seen, however, he continued to resist the idea that the self or subject is an entity. He also continued to deny that our relation to our mental states is one of possession or ownership. In the *Investigations* he writes that we are inclined to think that when we see a room, there is some object of perception, call it 'the visual room', that we have and no one else has (PI §§398–402/WR 196–7). The visual room is, as it were, a presentation on the stage of the theatre of representations. But, Wittgenstein argues, you cannot be thought to have or possess something which it is logically impossible for anyone else to have. The concept of possession makes no sense where a change of ownership is unintelligible. (This is not the only odd feature of the visual room. You are supposed to be more intimately acquainted with it than with the real room, yet the visual room is not such that you can even enter it.)

Wittgenstein suggests that there is no 'visual room' if that implies a kind of mental duplicate of the real room beyond the mind. There is just the room as you see it. To be a conscious being is not to possess a peculiar kind of inner mental world all one's own. (To portray consciousness in that way makes the mind-body problem particularly intractable: how could such a mental world be produced merely by goings-on in the brain? (PI §412/WR 200)).

These passages certainly contribute to Wittgenstein's critique of Cartesian conceptions of the mind. But if we are to abandon the solipsism of his early 'no-ownership' theory, yet maintain that the self is not an object and that mental states are not possible objects of possession, what kind of view of the self and its states is open to us?

Let us approach this question by considering Wittgenstein's mature view of the contrast between first- and third-person psychological ascription. Cartesian approaches afford an easy explanation of the fact that each of us is aware of his or her mental states in a way that no one else is: my mental states are objects that only I have, and the way in which I have them yields incorrigible knowledge of their character. What is Wittgenstein's alternative picture?

Wittgenstein does not take Gilbert Ryle's course and maintain that a person knows about her own mental states in the same way that she knows about other people's – by observing behaviour. Wittgenstein does not simply turn the mind inside out, making everything equally public; rather, he preserves a significant asymmetry between first- and third-personal relations to the mental. When I attribute mental states to other people I do so on the basis of behav-

journal criteria. This is not a process of speculating about something essentially hidden, for the minds of others are often manifest in their behaviour. What someone believes can be shown by his actions, his anger can be displayed by his expression, and so on. Nevertheless, our judgements about the minds of others are based on observation and other evidence, and there is room for error and misjudgement.³²

Wittgenstein portrays my relation to my own mind quite differently. I do not (usually) discover my own mental states by considering evidence or applying criteria. I neither observe the contents of my mind, nor infer what they are. Indeed, Wittgenstein paradoxically concludes that persons cannot properly be said to have knowledge of their own mental states at all. Knowledge implies a process of verification where error is possible, but no such process exists in most cases of first-person psychological ascription. Thus, when I see a red door, I cannot be said to know that I do, nor is it right to say that a person who is in pain knows that she is. Furthermore, when I state what I see or feel I do not describe the contents of my mind. First-person psychological ascriptions are not descriptions of events in a mental realm, but expressions or avowals. The statement, 'I am in pain' is a linguistically refined way of expressing or manifesting my pain. It is akin to a groan or moan.

Wittgenstein's view of first-person psychological utterances as avowals is a descendant of his earlier position in that it preserves the idea that the relation between thinking beings and their mental states is not to be understood as a relation between an entity, a self, and certain objects that it possesses. But the position departs from the solipsism of the earlier work by thinking of avowal as a mode of the manifestation of the mental in a public context.

Wittgenstein's view of avowals is often dismissed. First-person psychological utterances, after all, form the premises of inferences and figure in truth-functional operations. There is a logical symmetry between 'I am in pain' said by me, and 'Bakhurst is in pain' said by anyone else. It thus seems eccentric to try to associate such utterances with natural expressions that are grammatically inarticulate and logically inept.

Although Wittgenstein does sometimes exaggerate the difference between first- and third-person perspectives (as when he says, 'My own relation to my words is *wholly* different from other people's' (PI II 192/WR 206: my emphasis), I think his position might be developed to accommodate such objections. We have to concede that our mode of manifesting the mental in speech deploys expressions which have the *form* of descriptions. This allows them to stand in logical relations, and to capture the logical symmetry noted above, which makes possible the co-ordination of first- and third-person perspectives. Indeed, such utterances can be used as descriptions proper, in cases where, for example, we

do report our mental states (to a doctor, perhaps, or in confession). But it does not follow that the primary function of these expressions is not to express or manifest the mind. The basic idea that first-person psychological utterances are a mode of the articulation of the mental, while third-person psychological utterances aim to describe and report mental states, deserves further development.

The key question is whether such a position could be developed properly without taking 'I' to be a referring expression. If I am not a solipsist, I can admit that 'there is pain' and 'I am in pain' say the same thing only in cases where the pain is *mine*. Here the essential indexical element need not signify a mysterious form of ownership;³³ it is simply a component of a conceptual framework able to express the fact that others can be in pain just as I am. Moreover, it is important that what I can express is, not just that I have certain sensations like pain, but my conception of the world. First-person psychological ascription includes, of course, belief.³⁴ I can use first-person psychological ascription to manifest how I take the world to be. If a view of the world is to be seen as 'The world as I find it', we need the means to say that the world according to me is one among the ways the world might be found to be. I can recognize that a certain presentation of facts, or a certain narrative, is my way of finding the world only if I can see it as mine. To avow is to give voice, as one voice among others. The ineliminability of 'I' is a critical feature of any genuinely anti-solipsistic position. Self-effacement is impossible without self-reference. All this reinforces the idea that persons are the referents of 'I', for a person is precisely the sort of thing that has a conception of the world to express.

It might be replied that none of this shows that 'I' really *refers*. Why not say (following Glock paraphrasing Wittgenstein) that 'I' signifies the point of origin of the system of deixis, not a point on the deictic graph itself?³⁵ But how, we might ask, does 'I' signify any such thing? Surely, it does so by referring to a person. In any case, we should insist that the first-person perspective is more than a system of co-ordinates. It is, as it were, the site of a distinctive voice. It is hard to see how we could make sense of the expression of the mental, unless we understand such expressions as issuing from persons who are both in the world and originating sites of conceptions of the world.

7 Conclusion

The argument of the last three sections suggests the following. Wittgenstein's reflections on the immunity of 'I' to referential failure and error through misidentification do not succeed in overturning the *prima facie* case that 'I' is

a referring expression. Nor should we be concerned that if we grant that 'I' is a referring expression, we shall be forced to conclude that it refers to something like a Cartesian ego. We can hold that persons are the referents of 'I', and follow Strawson in our account of persons. Wittgenstein's reluctance to concede that 'I' refers can be traced back to the solipsistic framework of his early philosophy on which 'I' is held to be eliminable because redundant. The later doctrine of first-person psychological ascriptions as avowals departs from solipsism by portraying the mind as manifest in public space, but it is too undeveloped to yield a satisfying view of the first person. What we seek to avoid is the dualist framework in which, from the third-person perspective, selves appear as bodies, and when 'viewed from within' they shrink to vanishing points. The doctrine of avowals helps correct the former error, since it construes behaviour as imbued with mindedness, but it needs development if we are to escape the latter. When I think and speak of myself, I think and speak of something, a person, who is both of the world and conceiving of it. Moreover, my conception of the world includes myself (and not merely my body) as a participant in the world. Perhaps sense can be made of this without treating 'I' as a referring expression. I believe, however, that the intuitions that led Wittgenstein to deny that 'I' refers were an obstacle to his providing a truly compelling account of persons and their place in the world.

The shortcomings of Wittgenstein's views of the first person are especially disappointing because, as we noted at the outset, his remarks on rule-following and private languages intimate a radical view of human beings as embodied, socially and culturally situated beings, who attain the full conceptual capacities of persons through their appropriation of the human practices which constitute a form of life. Wittgenstein's remarks about the first person *per se* do little to enrich or develop this picture. Indeed, no positive vision of the self is allowed to emerge.

It might be thought that this complaint betrays a misunderstanding of the character of Wittgenstein's philosophy. His approach is therapeutic rather than constructive, aiming to exorcise philosophical confusion, rather than to build new theories about ourselves. It is difficult to see, however, how a purely therapeutic approach can be appropriate to problems of the self. When we consider philosophical scepticism about the 'external world', we have a clear sense of what we believe before philosophical mischief is made. Showing the illusions of scepticism returns us to a place where we find ourselves at home, secure in our ordinary conceptions of knowledge. But problems of the self are not like this. There is no pre-theoretical conception of ourselves in which to acquiesce when the intellectual fog has cleared. Philosophical, religious, mystical and popular-scientific pictures of our nature deeply infect our self-understandings, and no romantic vision of a return home is possible. Therapy

is thus not enough; philosophical reconstruction is essential. Wittgenstein's thoughts on the first-person are nevertheless a rich source of insight, and many later philosophers – Strawson, McDowell, Evans, Wiggins, among others – have made brilliant use of those insights in developing accounts, more ambitious than Wittgenstein's, of persons and their powers of self-consciousness. We are still a long way, however, from a satisfying philosophical presentation of the self and its place in the world.³⁶

Notes

- 1 It is an important question whether *only* persons are beings with a first-personal perspective. It can be argued that possession of such a perspective requires self-consciousness and mature conceptual capacities of a kind possessed only by persons, though this invites the reply that a first-personal perspective can hardly be denied to certain higher animals, human infants, and some disabled human beings, who cannot be considered persons in the full sense.
 - 2 Classic collections of writings on personal identity are J. Perry (ed.), *Personal Identity* (Berkeley, CA: University of California Press, 1975) and A. Rorty (ed.), *The Identities of Persons* (Berkeley, CA: University of California Press, 1976).
 - 3 John Locke, *An Essay Concerning Human Understanding*, ed. P. H. Niddich (Oxford, Oxford University Press, 1975), II.xxvii.6–29.
 - 4 Locke, *Essay*, II.xxvii.15.
 - 5 The history of the phenomenon of multiple personality disorder is traced by Ian Hacking in his *Rewriting the Soul: Multiple Personality and the Sciences of Memory* (Princeton, NJ: Princeton University Press, 1995).
 - 6 Classic discussions of this possibility are found in David Wiggins, *Identity and Spatio-temporal Continuity* (Oxford, Blackwell Publishers, 1967), §4.4, and Derek Parfit's *Reasons and Persons* (Oxford, Oxford University Press, 1984), ch. 12. Parfit's position, which is hostile to some traditional assumptions of the debate, is compared to Wittgenstein's below.
 - 7 Quoted from *American Psychologist*, 23 (1968), p. 724, by Thomas Nagel in 'Brain Bisection and the Unity of Consciousness', in Jonathan Glover (ed.), *The Philosophy of Mind* (Oxford, Oxford University Press, 1976), p. 120 (Nagel's paper was originally published in *Synthese*, 22 (1971), pp. 396–413).
- The evidence that 'two streams of consciousness' are exhibited consists of phenomena of the following kind. Since the right half of the visual field is seen with the left hemisphere, and vice versa, and each hemisphere controls one hand, if a different word is flashed in each half of the visual field (e.g. 'pencil' and 'toothbrush') and the subject told to find the corresponding object from beneath a screen, the hands will search independently, one retrieving the pencil while rejecting the toothbrush and vice versa (reported in Nagel, 'Brain Bisection', p. 115; he gives other examples).

My discussion of 'split-brain' cases is meant merely to motivate the thought experiment about the division of a person. I do not mean to imply that such a division is physically possible. Though we can separate the hemispheres, it is not likely that the lower brain can be divided in the same way. There may also be insuperable obstacles to the successful transplantation of a brain (see Parfit, *Reasons and Persons*, §87).

I should also note that some philosophers have recently challenged the significance of these split-brain cases; e.g. Daniel Dennett, *Consciousness Explained* (Boston, Little Brown, 1991), pp. 423–6.

8 Derek Parfit, *Reasons and Persons*, part 3, esp. chs 10–13.

9 See *Reasons and Persons*, pp. 236–43.

10 See *Reasons and Persons*, ch. 12.

11 See *Reasons and Persons*, p. x. Parfit comments on the relation between his position and Wittgenstein's on p. 273 of his book; though he correctly portrays what they have in common, his brief remarks do not do justice to the substantive differences that must issue from their contrasting methodologies.

12 *Consciousness Explained*, p. 422.

13 *Ibid.*, p. 423.

14 Constructivist views of persons are fashionable in various quarters, and Wittgenstein is sometimes cited in their support. Many such positions are based on views about classification or conceptualization, holding that something is a construction if it does not exist independently of our modes of classifying or conceptualizing it. When such views are applied to persons or selves it is important to distinguish three separate (though related) issues that are often mingled together: (1) Is the distinction between persons and other phenomena (e.g. things) a construct? (2) Are the criteria of personal identity constructs? (3) To what extent do the concepts we use to categorize people into kinds (concepts of race, gender, sexual orientation, sanity, intelligence, normality, etc.) pick out real or constructed properties? So, for example, Ian Hacking's Foucault-inspired view of 'making up people' is based primarily on an answer to (3), which in turn has implications for (2) (for instance, classifying certain forms of disturbed behaviour as 'multiple personality disorder' has implications for views of personal identity). I suspect, however, that Hacking would not claim that the distinction between persons and things is constituted by our modes of classification. See Hacking, 'Making Up People', in Thomas C. Heller, Morton Sosa and David E. Welby (eds), *Reconstructing Individualism* (Stanford, CA: Stanford University Press, 1986), and 'Why Multiple Personality Tells Us Nothing About the Self/Mind/Person/Subject/Soul/Consciousness', in David Bakhurst and Christine Synnott (eds), *The Social Self* (London, Sage, 1995), pp. 159–79.

Another variety of constructivism, espoused famously by the Russian psychologist Lev Vygotsky, though also endorsed by such philosophers as John McDowell and Charles Taylor, takes a genetic–developmental stance, arguing that individuals become persons only in virtue of their appropriation of conceptual powers through their inauguration into traditions, social practices, etc. Themes in

the later Wittgenstein, of course, also support such a view. See L. S. Vygotsky, *Thought and Language* (Cambridge, MA, MIT Press, 1986), Charles Taylor, *Sources of the Self* (Cambridge, Cambridge University Press, 1989), John McDowell, *Mind and World* (Cambridge, MA, Harvard University Press, 1994), esp. lecture V.

The two varieties of constructivism are not mutually exclusive. Some positions embody themes from both. See, for example, Jerome Bruner's *Acts of Meaning* (Cambridge, MA: Harvard University Press, 1990).

15 See Bernard Williams, 'The Self and the Future', in his *Problems of the Self* (Cambridge, Cambridge University Press, 1973), pp. 46–63.

16 We have to admit, however, that this may be a determinate answer which no one can know. Whether I continue to exist as a subject of experience in the relevant sense cannot be known from the third-person perspective, and if the operation has radically changed my psychological states, then I may have no sense of the continuity (or lack of it) either. One doesn't need much in the way of verificationist proclivities to find this consequence alarming, though cf. R. C. Swinburne, 'Personal Identity', *Proceedings of the Aristotelian Society*, 74 (1973–4), pp. 231–47.

17 Dennett, *Consciousness Explained*, p. 107.

18 In the selections chosen for *The Wittgenstein Reader* Wittgenstein does not explicitly deny that 'I' refers. He does so, however, in the unpublished manuscript MS 116, 215, and in his 'Notes for Lectures on "Private Experience" and "Sense-Data"' (LPE), PO 228.

19 This point is well made by Hans-Johann Glock and P. M. S. Hacker in 'Reference and the First Person Pronoun', *Language and Communication*, 16 (2), p. 100.

20 Evans takes a rather different view in *Varieties of Reference*, pp. 215–35.

21 See John Perry's interesting piece, 'The Problem of the Essential Indexical', in Quassim Cassam (ed.), *Self-Knowledge* (Oxford, Oxford University Press, 1994), pp. 167–83; first published in *Nous*, 13 (1), pp. 3–21.

22 It has been argued that if we treat 'I' as a referring expression, we are forced, absurdly, to say the same of, for example, the '-o' in the Latin *coquitur*: see Anthony Kenny, 'The First Person', in Cora Diamond and Jenny Teichman (eds), *Intention and Intentionality* (Ithaca, NY: Harvester Press, 1979), pp. 3–14. It does not seem absurd, however, to say that such verb-forms have a referential component determined by their conjugation (though word-parts have no specific referential function).

23 G. E. M. Anscombe, 'The First Person', in Cassam, *Self-Knowledge*, p. 152. Originally published in Samuel Guttenplan (ed.), *Mind and Language* (Oxford, Oxford University Press, 1974), pp. 45–65.

24 As Kant (in the third Paralogism) and, following him, Strawson have shown, no criteria of identity can be given for Cartesian egos. See Strawson's *The Bounds of Sense* (London, Methuen, 1966), part 3, ch. II, esp. pp. 168–9.

25 P. F. Strawson, *Individuals* (London, Methuen, 1959), ch. 3.

26 Glock and Hacker argue that treating 'I' as referring to the person who uses it allows us to explain the impossibility of reference-failure and error through misidentification in an unmythical, non-Cartesian manner – for if a speaker

utters 'I' intentionally, then she cannot be ignorant of whose utterance it is. Glock and Hacker prefer, however, to call 'I' a 'degenerate' referring expression (by analogy to Wittgenstein's claim that tautologies are degenerate propositions) in order to acknowledge that there are ways in which 'I' does not function like a paradigmatic referring expression. See 'Reference and the First Person Pronoun', pp. 98, 103.

27 See *Reasons and Persons*, pp. 210, 251.

28 Strawson, *Individuals*, p. 97.

29 For sustained critiques of Parfit's reductionism, see John Campbell, *Past, Space, and Self* (Cambridge, MA, MIT Press, 1994), ch. 5, and John McDowell, 'Reductionism and the First Person', in Jonathan Dancy, *Reading Parfit* (Oxford, Blackwell Publishers, 1997), pp. 230–50.

30 David Hume, *A Treatise of Human Nature*, 2nd edn, ed. L. M. Selby-Bigge, revised P. H. Niddich (Oxford, Oxford University Press, 1978), I.iv.vi.

31 People sometimes write of 'no ownership' or 'no-subject' conceptions of the self (e.g. Strawson, *Individuals*, p. 93). There is an irony here that can be confusing, since on these views there is no self for there to be a conception of.

32 It is common in contemporary philosophy of mind to represent knowledge of the mental states of others as a form of theoretical knowledge. Our vocabulary for psychological description is portrayed as comprising a 'folk theory' of mind and developmental psychologists ponder the question of when and how a child acquires her 'theory of mind'. Although Wittgenstein stresses that our judgements of the mental states of others are verified by appeal to certain sorts of evidence, he does not treat knowledge of other minds as theoretical in the contemporary sense. For Wittgenstein, it is neither the case that our knowledge of other minds is 'merely' theoretical knowledge about something that could be known more directly, nor is it the case that minds are theoretical constructs. Wittgenstein also denies that the question of whether other beings have minds at all is a matter of theoretical speculation. I do not form an opinion that others have minds. My conviction that they do is evinced in my attitude to them, an attitude that can be suspended only in extreme or artificial circumstances (PI §420).

33 We should perhaps hold, against Wittgenstein, that persons do possess (or at least 'have') mental states, but that there is nothing mysterious about this once Cartesian misconceptions have been dispelled. As I noted at the end of §5, Strawson argues that some such concept of possession enters into the very conditions for the identification of mental states, since we cannot identify them except as the states of some particular person. A Wittgensteinian who has outgrown Lichtenbergian ideas ought not to deny this, especially if she wishes to resist Parfit's reductionism.

34 Although much of Wittgenstein's treatment of avowals focuses on pain and other sensations, he expresses a similar view of beliefs (and, presumably, other propositional attitudes). This emerges in his discussion of Moore's paradox (PI II, 190–2/WR 204–7). Moore noted that it is absurd to claim, 'It is raining, but I do not believe it'. Wittgenstein took the paradox to show that the role of 'I' believe

that *p* is not to describe how things are with the speaker but to express or avow a conviction. This is also the role of simply asserting *p*, hence the inconsistency. Hans-Johann Glock, *A Wittgenstein Dictionary* (Oxford, Blackwell Publishers, 1996), p. 163.

36 I am grateful to the Social Sciences and Humanities Research Council of Canada and to the Principal's Development Fund of Queen's University for supporting the research that led to the writing of this essay. I am indebted to participants at the Queen's Philosophy Colloquium and to members of Reading University's Philosophy Society, and especially to Jonathan Dancy and Brad Hooker, for their helpful discussion of an earlier draft. I must also thank Hanna Pickard and Christine Sympson for their insightful criticisms.