The Blackwell Guide to Metaphysics

Edited by

Richard M. Gale

Blackwell Publishers

Editorial Offices:
108 Cowley Road, Oxford OX4 1JF, UK
Tel: +44 (0)1865 791100
350 Main Street, Malden, MA 02148-5018, USA
Tel: +1 781 388 8250

First published 2002 by Blackwell Publishers Ltd

Library of Congress Cataloging-in-Publication Data has been applied for.

ISBN 0-631-22120-4 (hardback); ISBN 0-631-22121-2 (paperback)

A catalogue record for this title is available from the British Library.

Set in 10 on 13 pt Galliard
by SNP Best-set Typesetter Ltd., Hong Kong
Printed and bound in Great Britain
by T.J. International, Padstow, Cornwall

For further information on
Blackwell Publishers, visit our website:
www.blackwellpublishers.co.uk

The Blackwell Guide to Metaphysics
Edited by Richard M. Gale

Chapter 9

What Do We Refer to When We Say "I"?

Peter van Inwagen

I will begin by asking you to consider certain words and phrases whose meanings are obviously closely related – closely enough that you will see what I mean if I say that these words constitute a family: 'soul', 'self', 'person', 'ego', 'I' (used as if it were a common noun, as when Descartes refers to 'this I'), 'mind' (used with the implication that the things it refers to are *objects*, substances in the metaphysical sense of 'substance'). I think you will agree that the meanings of these words are indeed closely related. Perhaps you will also agree that it is not always entirely clear what these words do mean, or how their meanings are related. Questions about the meanings of and the relations between the meanings of the words in this family are, in my view, best framed in terms of their relations to 'I' – the first-person singular pronoun, that is, not the pseudo-noun. Thus, for example, we can explain the difference between St. Thomas's and Descartes' use of 'mind' and 'soul' (*mens* and *anima*) by pointing out that Thomas did not think that when he used the word 'I' (or 'ego' or whatever) he referred to his mind or his soul, and Descartes thought that when he used the word 'I' he referred to both his mind and his soul. Or here is an autobiographical example: whenever I hear present-day philosophers going on about "selves" – asserting, perhaps that modern neurobiology has exploded the old myth of the self or that the self is a social construct or that Descartes was mistaken in thinking that a sharp boundary could be drawn between self and world – the first thing that I always ask these philosophers is whether, when I use the word 'I' I refer, or at least am attempting to refer, to one of these "selves" (my own, of course). After all, if there are selves and if, when I use the word 'I' I refer to something, it would seem that it must be my Self I refer to.[1] Or if there is such a thing as my Self, and I do *not* refer to it when I use the word 'I', how could it be correct to call this thing my Self? It is not I, it is rather something numerically distinct from me, and how can something that is not I be properly called my Self? Or, if the philosophers I am talking to are of the party that holds that selves are myths, I ask them whether their position is that they do not exist – for if they exist, then, of course, each time one of them uses the word

'I', that use refers to something, and what could that referent be but the self of the speaker? These questions may seem to some to be trivial quibbles on my part, but they are no such thing. They confront the philosophers who talk of selves with a dilemma I have never seen satisfactorily resolved. If they say, "Yes, that's just what your Self is (or that's just what it would be if there were such a thing): what you refer to when you say 'I'," then their theses almost invariably turn out to be nonsense or obviously false or so obviously true that it is hard to think why anyone would bother stating them. (Modern neurobiology has obviously not shown that there are no such things as you and I.) Or, if they say, "No, that's not what your Self is – your Self is not you but something numerically distinct from you; it is [or 'is supposed to be'] something you *have*; it's not what you *are*," then they are never able to give any real explanation of what they mean by 'self': their attempts at explanation turn out to be so much semantical arm-waving.

Well, then, what *do* we refer to when we say 'I'? I am sorry to say that there seem to be nine possibilities. I begin with this one:

(1) We refer to nothing.

Many philosophers have endorsed this position. The endorsements are mostly of two sorts: the old-fashioned "Humean" sort, or the more modern "Wittgensteinian" sort.[2] Hume, or so I interpret him, held that there is just nothing *there* for the word 'I' to refer to. If there were, we should be able to find it in introspection, and we find no suitable referent for the word when we enter most intimately into what we call *ourselves*. What we find in introspection are impressions and ideas that would be qualities of the referent of 'I' if it had one; but since (we find) there is nothing "in there" to be the referent of the word, there are only the impressions and ideas, free-floating qualities that inhere in no underlying substance. One who took the general Humean line might of course say that the word 'I' referred to some *collection* of these qualities, but collections of ideas aren't really suitable candidates for the referent of 'I' (or so it might be argued) because it is part of the meaning of the word 'I' that its referent is something that persists through changes of qualities, and that is just what collections of qualities don't do. The Wittgensteinian view, most clearly stated in Elizabeth Anscombe's well-known essay "The First Person,"[3] is that it is not the function of the word 'I' to refer; the word is thus unlike "the present king of France," which is in the denoting business but is a failure at it; rather, the word, despite the fact that it can be the subject of a verb or (usually in its objective-case guise, 'me') the object of a verb, is not in the denoting business at all. Thus, for Hume, the word 'I' refers to nothing in the way 'the present king of France' refers to nothing; for Professor Anscombe, the word 'I' refers to nothing in a way more like the way in which 'if' and 'however' refer to nothing.[4]

The remaining eight possibilities – all, of course, cases of "We refer to something' – are generated by the possible ways of picking one each from the pairs

'transitory'/'lasting', 'enduring'/'temporally extended',[5] and 'material'/'immaterial'. ('Physical' and 'natural' might be alternative readings for 'material'.) They are:

(2) We refer to something transitory and enduring and material.
(3) We refer to something transitory and enduring and immaterial.
(4) We refer to something lasting and enduring and material.
(5) We refer to something lasting and enduring and immaterial.
(6) We refer to something transitory and temporally extended and material.
(7) We refer to something transitory and temporally extended and immaterial.
(8) We refer to something lasting and temporally extended and material.
(9) We refer to something lasting and temporally extended and immaterial.

The most common answers to the question "What do we refer to when we say 'I'?" are special cases of the general possibilities I have labeled (4), (5), (6), and (8). Some examples would be:

(4) Many materialists, those who accept an "endurantist" or "three-dimensionalist" theory of identity across time.

(5) Most idealists (Berkeleyan, not Absolute) and dualists. (All or almost all idealists and dualists are endurantists; it may be that Jonathan Edwards was a dualist and a "temporal extentionalist" – a lonely exemplar of possibility (9).)

(6) Many materialists, those who accept a "perdurantist" or "four-dimensionalist" theory of identity across time and who hold that an utterance of the word 'I' at the time *t* denotes a "time-slice" of the utterer, the slice taken at the time *t*. (These are the philosophers who hold that phrases like 'Peter-now' and 'Peter-at-noon-yesterday' are denoting phrases and that they denote numerically distinct objects, objects related by "gen-identity" rather than identity.)[6]

(8) Many materialists, those who accept a "perdurantist" or "four-dimensionalist" theory of identity across time and who hold that an utterance of the word 'I' at the time *t* does not denote the time-slice of the utterer taken at the time *t*, but denotes rather the "whole four-dimensional individual," the mereological sum of all the time-slices related to *t*-slice by gen-identity.

My purpose in this essay is not to endorse any one of these positions – I am in fact an adherent of (4) – but to try to show something that seems to me to be important about the two very popular positions (4) and (5): they cannot be coherently combined with the psychological-continuity theory of personal identity. I will argue for the following two conclusions: that any materialist who accepts a psychological-continuity theory of personal identity must accept not (4) but (8); that any immaterialist (any dualist or idealist) who accepts a psychological-

continuity theory of personal identity must accept not (5) but (9). The proponent of a psychological-continuity theory of personal identity, in other words, must be a perdurantist (or temporal extensionalist) and not an endurantist.

Let us begin by considering a dualist who accepts a psychological-continuity theory of personal identity. Let us consider John Locke. Locke believes that when I utter the word 'I', I refer to my soul, to an immaterial substance. He also accepts – as untold generations of philosophy students have been informed in one of the first philosophy lectures they have attended – a "memory" criterion of personal identity.[7] (A memory criterion of personal identity is, of course, a species of psychological-continuity criterion of personal identity.) Now suppose that in 1990 all my memories were obliterated – that my soul became once more the *tabula rasa* that, in Locke's view, she was at the beginning of her existence. And let us suppose that experience immediately began once more to "write" on the tablet of my soul, or rather the soul that was mine before 1990, and that presently, owing to this influx of useful information, this soul once more became capable of ratiocination and (being still properly connected with the vocal apparatus that had once been mine) speech. Then she, or the man whose soul she is, is once more capable of producing meaningful utterances of the word 'I' and, when she does produce them, they refer to the soul that once was, but is no longer, mine. Let us distinguish "pre-traumatic utterances of 'I' that proceeded from *this* vocal apparatus" and "post-traumatic utterances of 'I' that proceed from *this* vocal apparatus" – the "trauma" being the conversion in 1990 of what was till then my soul to a *tabula rasa*. And let us give the soul that was mine till 1990 the proper name 'Anima'. It is clear that Locke's philosophy of personal identity entails all three of the following propositions:

> The referent of the pre-traumatic utterances of 'I' that proceeded from this vocal apparatus = Anima
>
> Anima = the referent of the post-traumatic utterances of 'I' that proceed from this vocal apparatus
>
> The referent of the pre-traumatic utterances of 'I' that proceeded from this vocal apparatus ≠ the referent of the post-traumatic utterances of 'I' that proceed from this vocal apparatus.

(The third proposition is entailed by the memory criterion of personal identity; if this proposition were false, then the post-traumatic utterer of 'I' could say, and say truly, "I existed before the trauma" – and this he cannot do if the memory criterion is correct, since, by definition, he has no memories of anything that preceded the trauma.) But to assert all three of these propositions is obviously to fall into logical incoherency, for they together constitute a violation of the principle of the transitivity of identity – and hence, a violation of the principle of the indiscernibility of identicals, of which the transitivity of identity is an immediate consequence. And how does Locke fall into this incoherency? Obviously as a result

of accepting the memory criterion of personal identity, for it is that principle that has the consequence that the person (myself) who called Anima 'I' before 1990 is not the person who later called Anima 'I'.

If this argument is too complicated for your taste, here is a simpler one. Suppose that when I utter the word 'I' I refer to Anima. Then I *am* Anima – for the same reason that if, when I utter "the largest structure in Egypt" I refer to the Great Pyramid, then the largest structure in Egypt *is* the Great Pyramid. That is how reference works. And if I am Anima, then I am logically stuck with being Anima – and Anima is logically stuck with being me, for the plain reason that a thing and itself cannot go their separate ways. And, therefore, Anima is always going to be me (as long as she exists, anyway) no matter what happens to her. If all her memories are obliterated, that will no doubt be a grave misfortune for her, or for the man whose soul she is, but it won't turn her into something or someone else. The thing about logical truisms is, there is just no way round them, and the following is a logical truism: no misfortune, however grave, can turn someone into someone else, for nothing can turn someone into someone else. But the memory criterion of personal identity has the consequence that Anima can be me at one time and someone else at a later time.

The logical incoherency of Locke's position has nothing in particular to do with his belief that when one uses that word 'I', one refers to an immaterial soul. Plenty of materialists have fallen into exactly the same incoherency. If the materialists are right, and if, when I use the word 'I' I refer to something, then I refer to something material – for the only alternative is that I refer to something immaterial, and if I referred to something immaterial, there would *be* something immaterial and materialism would be false. But plenty of materialists believe in the conceptual (if not the technological) possibility of a certain sort of "bodily transfer," and it is these materialists who have fallen into the same incoherency as Locke. Sydney Shoemaker is a good example of a materialist who believes in the possibility of this sort of bodily transfer, or at least takes its possibility very seriously.[8] Shoemaker, although he is a materialist, holds that it is possible for a person to "change bodies" – or at least he holds that there are good reasons to think that bodily transfer is possible, even if these reasons are not absolutely conclusive. And he does not think that changing bodies requires a "brain transplant" or any other procedure that involves moving matter from one human body to another. In his view, it is entirely plausible to suppose that (even if it is not self-evident that) a transfer of the information contained in my brain to a suitably prepared "bland" brain in another human body would suffice for my acquiring a new body – at least if my "original" brain is destroyed or turned into a "blank" in the process. (In the sequel, I will for convenience's sake write as if Shoemaker accepted without qualification the possibility of bodily transfer simply in virtue of a flow of information.)

If we use the common noun 'person' for those things that are referred to by uses of the personal pronouns ('I', in particular), Shoemaker's position is that it is possible for a person (a material thing) to change bodies; Locke's position was that it was possible for a person (an immaterial thing) to change souls. An

argument exactly parallel to the argument I used to show that Locke's position was incoherent can be used to show that Shoemaker's position is incoherent.[9] Here is the simple version. Let 'Hylas' be the material thing I refer to when I use the word 'I'. (There must be such a thing if I refer to something when I use the word 'I' and if – as the materialist contends – everything is material. That's logic, as Tweedledee said.) Then I *am* Hylas – for the same reason that if when I utter "the tallest structure in Paris" I refer to the Eiffel Tower, then the tallest structure in Paris *is* the Eiffel Tower. That is how reference works. And if I am Hylas, then I am logically stuck with being Hylas – and Hylas is logically stuck with being me, for the plain reason that a thing and itself cannot go their separate ways. And, therefore, Hylas is always going to be me (as long as he exists, anyway) no matter what happens to him. If all Hylas's memories – *my* memories – are obliterated and their informational content somehow transferred to and caused to be embodied in some appropriately structured but numerically distinct material thing *x*, that will not cause Hylas to become *x*. The thing about logical truisms is, there is just no way round them, and the following is a logical truism: no transfer of information, however perfect, can turn a thing and another thing into a thing and itself, for nothing can turn a thing and another thing into a thing and itself. (Hylas and *x* are a thing and another thing; if Hylas became *x*, Hylas and *x* would be a thing and itself: that is, there would be only one of them. Identity is, after all, identity; it is what it is, and not some other relation.) Bodily transfer by a flow of information is therefore impossible.

It is important to note that in this argument 'Hylas' does not necessarily refer to what is commonly called 'my body' – to a "whole" human organism. Rather, 'Hylas' refers to *whatever* material thing it is that I am. Other possible candidates – other than what is commonly called my body – for the referent of 'Hylas' would be: my brain and central nervous system (which Sellars has called the "core person"), my brain, whichever of my cerebral hemispheres it is that controls my use of language and thus is the source of all those occurrences of the word 'I' that you have been exposed to in this essay (this is the position of Roland Puccetti), my cerebral cortex (commonly supposed to be the seat of conscious experience), my pineal gland (so might Descartes have said in the unlikely event of his conversion to materialism), and a tiny material particle that, although it is probably located somewhere in my brain, has so far eluded the observations of brain-physiologists (R. M. Chisholm once held this view[10]). This argument, therefore, does not assume that the materialist is committed to the premise that I am identical with what is commonly called my body; it assumes only that the materialist – the materialist who does not deny that I and other persons exist – is committed to the thesis that I am identical with *some* material thing. I said that my conclusion was that bodily transfer (in virtue of a flow of information) was impossible. But this way of formulating my conclusion captures its whole content only if – on the assumption that human persons are material things – one's "body" is whatever material thing one is identical with. On this understanding of the word 'body', if I am my pineal gland, then my body is a small pine-cone-shaped outgrowth of

my forebrain, and not the whole human organism inside which this little structure makes its home.

Shoemaker has recently tried to show that my argument for the conclusion that (given the assumptions I have made) a person cannot change bodies simply in virtue of a flow of information is mistaken.[11] The mistake, he says, consists in my supposing that those who believe in the real existence of persons – who believe that when one uses the pronoun 'I' one really does refer to something – are committed thereby to the position that persons are *individual substances*, that they are what he calls "(relatively) autonomous self-perpetuators," things that persist through time (at least largely) in virtue of ongoing internal processes or "immanent causation." Consider, by way of contrast, the Privy Council of an autocratic monarch, a body whose continued existence and whose membership at a given time depend on and only on the decree of the monarch. If the Privy Council really exists, it is a good example of a thing that is *not* an autonomous self-perpetuator, since its continued existence and its membership at a time depend entirely on things outside itself.[12] (It is thus unlike a private club, which can gain new members only by the actions of those who are already its members.) And, if we are materialists and believe the Privy Council really exists, we must believe that the Privy Council is a material thing. Suppose that Elizabeth – our autocratic monarch – declares to the Privy Council, assembled in London at noon, January 1, 1590, "I'm giving *you* all the sack. I hereby appoint the following persons to this council." She proceeds to recite the names of ten men all of whom happen to be in York at the moment. Then the Privy Council is translated instantly to York. Despite its being a material object, it manages this translation without ever occupying any point in space between London and York. This translation, it will be noted, does not require even a transfer of information. If we supposed, however, that a person could become a member of the Privy Council only by accepting the offer of an appointment to it, a transfer of information from London to York would be necessary for the translation; but *only* a transfer of information would be necessary, and even in 1590 it was possible to transfer information from London to York without causing any material thing to move from one city to the other. Thus, if Shoemaker is right, there is no *logical* barrier to the translation of a material thing from one place to another simply in virtue of a transfer of information between the places. All that is necessary is that the translated material thing *not* be a substance, an autonomous self-perpetuator, a thing whose identity across time depends on immanent causation.

That the instantaneous translation from London to York of a material thing is a feature of our imaginary case follows simply from the premises that the Privy Council really exists, that *it* is at one moment in London and a moment later in York, and that everything is material. And, Shoemaker argues, since we have strong intuitions that favor the thesis that a perfect transfer of the information in one brain to another brain would (at least under certain conditions) be "person-preserving" – and, more generally, strong intuitions that favor a psychological-continuity criterion of personal identity – we have a strong motivation for

believing that a person can change bodies merely in virtue of a transfer of information from one body to the other. (Locke, of course, could offer essentially the same argument for the conclusion that we have a strong motivation for believing that a properly conducted transfer of information from one soul to another would result in a person's changing souls, and that this belief faces no logical difficulties.)

Does the "Privy Council" example show that it is possible for a material thing to change places simply in virtue of a flow of information between those places? I think we can see that it does not if we ask ourselves this question: What material object is the Privy Council? There are, we know, twenty men (men, we are assuming, are material objects), ten in London and ten in York, who, at various moments, in some sense make up or constitute the Privy Council. But what is this "making up" or "constitution"? What relation do these men bear to the Privy Council? I can't see any relation for this relation to be but that of part to whole. That is: if τ is the moment of the supposed translation of the Privy Council from London to York (noon, January 1, 1590), then, immediately before τ the Privy Council is the mereological sum of ten men in London, and immediately after τ the Privy Council is the mereological sum of ten men – ten *other* men – in York. (Mereological summation is defined as follows:

> *x* is a mereological sum of the *y*s at $t =_{df}$
>
> At *t*, all the *y*s are parts of *x*, and everything that is a part of *x* at *t* then overlaps [shares a part with] one of the *y*s.

The mereological sum of the *y*s at *t* is the unique object that is a mereological sum of the *y*s at *t*.)

One might object that it is a rather naive social ontology that identifies a social entity like a council, team, corporation, or sect with the mereological sum of its members. And I would agree: that is to say, I would agree that it is a rather naive social ontology that maintains that (given that individual human beings are material objects) a social entity is a material object. No doubt it is a much more plausible thesis that a social entity is some sort of "logical construct." (To say that General Motors is a logical construct is to say either that 'General Motors' does not denote anything and that the true sentences in which this term occurs can be paraphrased as sentences in which it does not occur, or else to say that General Motors is some sort of set or other abstract object.) But it is essential to the point of the example that the Privy Council be a material object, and if it is a material object, there doesn't seem to be any material object for it to be but the mereological sum of its members. If the example is to provide a case of a material object that is translated from London to York simply in virtue of a flow of information, then the following must be true: before τ the Privy Council is the mereological sum of ten men in London, and after τ it is the mereological sum of ten men in York.

Now suppose that immediately before τ, someone in London had said, "See those ten men there? I hereby name their mereological sum 'Londinium'." And

suppose that immediately after τ, someone in York had said, "See those ten men there? I hereby name their mereological sum 'Eboracum'." Can it be that Londinium *was* Eboracum? – that 'Londinium' and 'Eboracum' are two names for one thing? If the Privy Council example is to be an example of a material object changing its position simply in virtue of a flow of information between two places, this will have to be the case. Here is a consecutive account of the sequence of events in our story. Londinium was sitting there in London. Elizabeth spoke a few words. Londinium instantly lost all its proper parts and, without having moved, found itself in York with a new set of proper parts – whereupon someone conferred the new name 'Eboracum' on it. And some other strange things may have happened as well. Consider the ten men in York. If, immediately before τ, they had a mereological sum, this object was either annihilated at τ or at least changed some of its parts – it was immediately after τ composed of some five of the ten men who had composed it a moment before, or it was composed of the parts that had a moment before composed York Minster, or something of that general sort. And let's not forget the ten men in London. If immediately after τ they had a mereological sum, either this object was created *ex nihilo* at τ, or else it had before τ a different set of parts, some or all of which it instantly discarded as a necessary concomitant of becoming the mereological sum of those ten men. (I have been assuming that for any *x*s, those *x*s have at most one mereological sum at a given time. Other assumptions are possible – possible in the sense that they are not ruled out by the definition of a mereological sum. Suppose that just before τ, each set of ten men had six mereological sums. Perhaps only one of them, whichever one it was that was the Privy Council, was translated: after the translation, the ten men in London had only five mereological sums and the ten men in York had seven.) And, remember, all these things happened because an irascible queen spoke a few words. If she hadn't said, "I'm giving *you* all the sack. I hereby appoint the following persons to this council . . . ," Londinium would have remained in London and would have continued to be composed of the same ten men.

That this could happen looks to me like an excellent candidate for an incoherent thesis. I concede that I can't derive a formal contradiction from it without introducing a premise that some might dispute. (Any of the following three premises would do: that an object can't instantaneously lose all its proper parts and continue to exist; that if certain objects have a mereological sum at two different times, then their sum at the one time is identical with their sum at the other; that the identity of the mereological sum of a given set of objects can't be determined by a decree, even a royal one.) But, I would ask, is the thesis that Londinium changed position instantaneously a better candidate for ontological coherency than the following thesis: The Prime Minister changed position instantaneously when "he" switched from being John Major to being Tony Blair? Given that Privy Councils are mereological sums of their members, isn't *this* what Elizabeth's decree would cause to happen: Londinium stays in London and continues to be the sum of the same ten men; Eboracum was in York before the decree

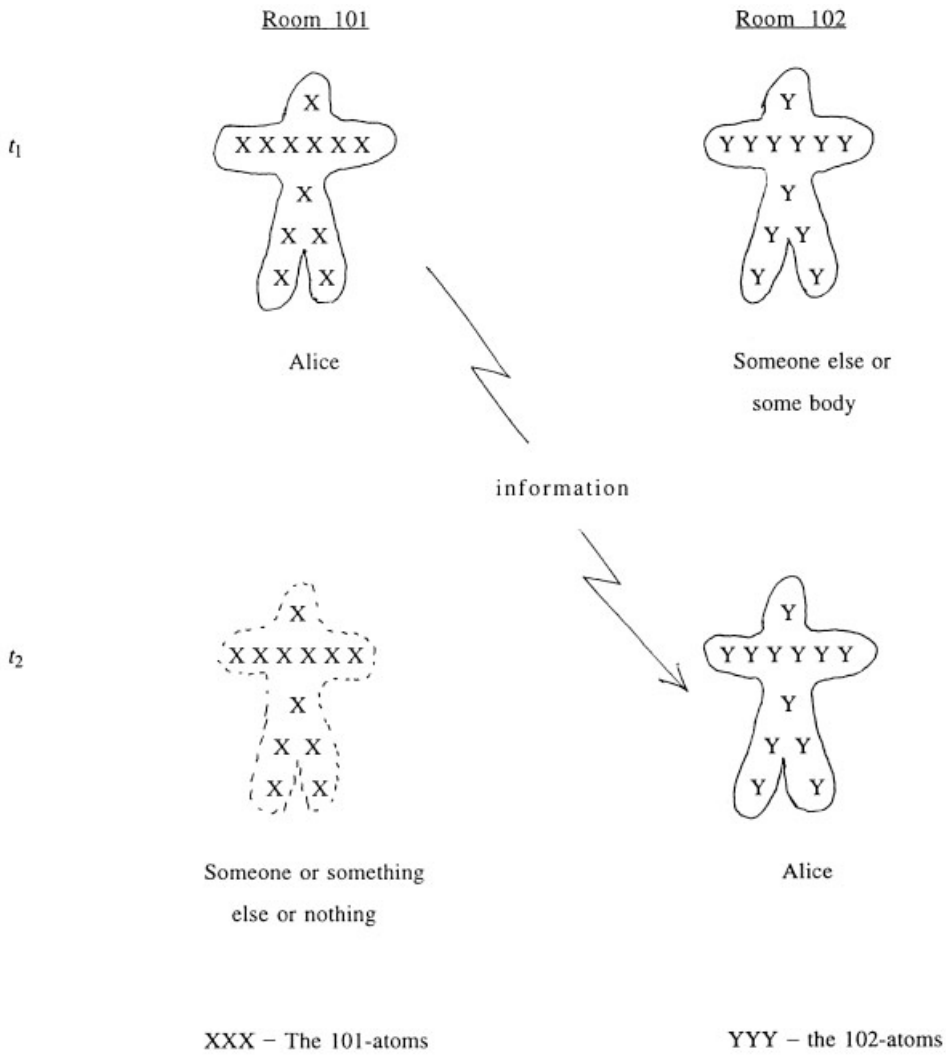

Figure 9.1

and remains in York and continues to be the sum of the same ten men; the title "the Privy Council" is transferred from Londinium to Eboracum?

We can apply essentially the same argument "directly" to Shoemaker-style body-changes. (This application of the argument is illustrated in figure 9.1.)

Suppose we intend to "transfer" our friend Alice "to another body." If Alice really exists and is a material thing, she is now (at t_1) the mereological sum of certain atoms. Here is what would have to take place if we successfully transferred her to another body. The atoms whose sum she is are now in Room 101. (They are represented by Xs in figure 9.1.) Certain *other* atoms (represented by Ys), atoms to be found in Room 102, compose (now, at t_1) some other human being or some human body other than hers. Information and nothing else passes from Room

101 to Room 102 (or "nothing else" besides whatever must, of metaphysical necessity, move from Room 101 to Room 102 if information flows in that way). Solely in virtue of this flow of information, the object that had been the mereological sum of the atoms in Room 101 becomes (at t_2, almost immediately after t_1) the mereological sum of the atoms in Room 102. The atoms that *had* composed that object, the atoms in Room 101, either cease to have a mereological sum or immediately acquire a new mereological sum, and the object that had been the mereological sum of the atoms in Room 102 is no longer the mereological sum of those atoms – either it is destroyed or it becomes the mereological sum of some other atoms.[13] (In the diagram, a solid outline around a group of atoms represents those atoms as having a mereological sum. A dotted outline around a group of atoms represents our declining to take a stand on whether those atoms have a mereological sum.) Well, you can say this and I can't catch you in a formal contradiction – unless I help myself to some premises that you might want to reject. But can you really suppose that your position is coherent? Isn't *this* what would really happen when the machinery was put into operation: Alice stays in Room 101 – or else she is destroyed, depending on what is done with the atoms in Room 101 – and some unfortunate woman in Room 102 is turned into a psychological duplicate of Alice? That is, wouldn't things happen in the way illustrated by *this* diagram (figure 9.2)?

Shoemaker's position is therefore incoherent. At least it has some very odd consequences, consequences that seem to *me* to be incoherent. We may ask Shoemaker to respond to the following dilemma. Consider the story of Alice. In this story, either some material thing that was in Room 101 when the story began was in Room 102 when the story ended, or else no material thing that was in Room 101 when the story began was in Room 102 when the story ended. In the latter case, materialism is false, since Alice was in Room 101 when the story began and in Room 102 when the story ended. But if we say that some material thing that was in Room 101 when the story opened was in Room 102 at the close of the story, we seem to have endorsed the possibility of a kind of "movement" comparable to the movement of the Prime Minister when he changed from being Major to being Blair – which is at the very least an excellent candidate for incoherency.

Now it might be objected that the above arguments, even if they are completely successful, show only that Position (4) is inconsistent with the possibility of bodily transfer (*sc.*, by flow of information) and not, as promised, with the psychological-continuity theory of personal identity; for we have not shown that the psychological-continuity theory entails the possibility of bodily transfer. Here I will simply assume that it would be at least very odd for the proponent of the psychological-continuity theory to reject the possibility of bodily transfer: why *couldn't* the psychological states tokened in one body be continuous with those tokened in another body?[14]

But if you are a friend of body-change operations, do not despair. One can have body-change operations if one does not make the assumption that persons endure

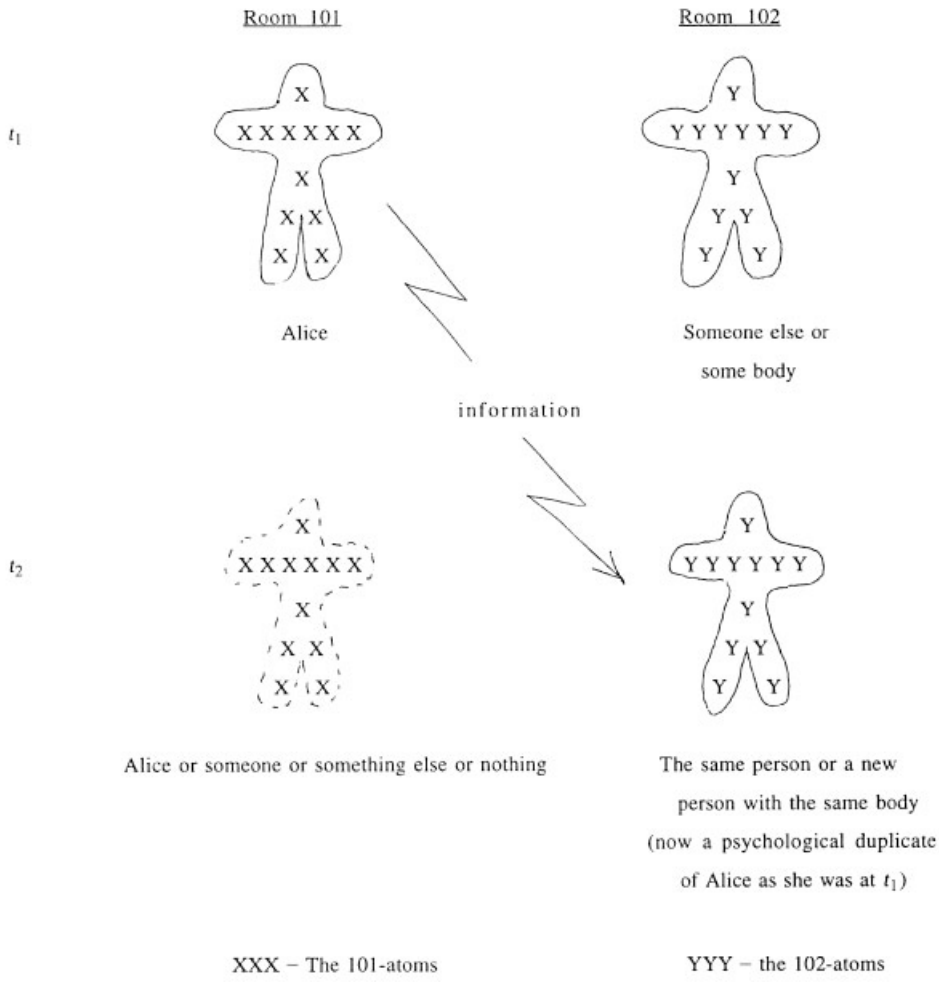

Figure 9.2

through time – that is, if one is willing to make the assumption that persons are extended in time. This is the position of David Lewis, who has applied it to questions about the nature of the human person and the identity of the person across time with his usual technical perfection.[15] The essential trick is this:

> Material objects are four-dimensional things, things extended in time as well as in space: what we normally think of as cases of objects that "endure through time" are actually cases of objects that are extended in time. Any two such four-dimensional objects have a mereological sum that is itself a four-dimensional object. Certain four-dimensional objects count as persons. A four-dimensional object is a person if it is a maximal aggregate of person-stages – a person-stage being a four-dimensional object that would be a person if it existed "all by itself." Leave aside the question of the meaning and purpose of the qualification "maximal." A mereological sum of person-

stages is an "aggregate" if the stages are psychologically continuous with one another in the right sort of way.

Given this view, the outcome of a successful "bodily transfer" between Room 101 and Room 102 may be described as follows. Alice is, like all of us, a four-dimensional object, a maximal aggregate of person-stages. Unlike most of us, however, she is not a spatially continuous four-dimensional object. She is, rather, the sum of two individually spatially continuous aggregates of person-stages that are not spatially connected with each other. The earlier of the two ends in Room 101, and the later begins in Room 102 almost immediately afterwards. Despite the fact that these two aggregates are not spatially connected, they are (owing to the operations of the "bodily transfer machine") psychologically connected, and in the right sort of way for the two aggregates together to form a maximal aggregate of person-stages – that is, a single person.

But to accept the theory of personal identity on which this story is based is to reject position (4) in favor of position (8) – or, if one is a dualist like Locke, to reject (5) in favor of (9): to become a temporal extentionalist. (And it is not simply to become a temporal extentionalist with respect to persons, but with respect to everything temporal. After all, it could hardly be that although *some* material objects, persons, are extended in time, all other material objects endure through time.)

My conclusion is that (4), (5), (8), and (9) are all at least initially viable theories of the nature of the referent of 'I'. Nevertheless, anyone who accepts the possibility of bodily transfer – anyone, in fact, who accepts any sort of psychological-continuity theory of identity across time – cannot accept (4) or (5). That philosopher must become a temporal extentionalist.

Notes

1 Whenever I follow a possessive pronoun like 'my' or 'your' by the word 'self', I will capitalize 'Self' – just to make it clear to the reader that I am not writing 'myself' or 'yourself'.

2 According to Hume and the Wittgensteinians, 'I' refers to nothing because of considerations peculiar to the self or the first person. Other philosophers would endorse this position as a consequence of some very general metaphysical view, one that entails that *all* those things that are normally thought of as individual things are in some sense unreal: Parmenides, Spinoza, the Absolute Idealists, the adherents of certain Eastern religions, Bertrand Russell (at some points in his career), Peter Unger (at some points in his career).

3 G. E. M. Anscombe, *Metaphysics and the Philosophy of Mind: Collected Philosophical Papers, Volume II* (Minneapolis: University of Minnesota Press, 1981), pp. 21–36.

4 This comparison is mine and not Anscombe's. It has an important weakness: 'if' and 'however' do not occur in nominal positions, and thus no one is even tempted to regard them as having referents.

5 An enduring object is one that, well, "endures through time"; a temporally extended object is one that is extended in time in a way analogous to the way in which ordinary material objects are extended in space. In an earlier version of this essay, I used the terms 'three-dimensional' and 'four-dimensional' instead of 'enduring' and 'temporally extended'. Richard Swinburne pointed out to me that applying the former pair of terms to an immaterial soul implies that the soul is extended in space, which can hardly be an accurate representation of the views of those who believe in immaterial souls. This was, as Jonathan Bennett likes to say, a fair cop.

6 At least these philosophers hold that the referents of utterances of 'I' are four-dimensional objects if they admit that these referents have to have *some* extension in time, to be "time-slices" that, like slices of bread, have some thickness. It is hard to see how the utterance of an indexical word like 'I' could pick out a time-slice of zero temporal extent – just as it is hard to see how an utterance of 'here' could pick out a dimensionless point in space.

7 Paul Helm has suggested to me that Locke did not hold a "memory criterion of personal identity" – he held rather that the deliverances of memory constitute the primary evidence that we appeal to when questions of personal identity are in dispute. I am willing to grant that there are passages in Locke that support this interpretation; but Locke does sometimes at least talk as if memory *constituted* personal identity. The famous §10 of the chapter "Of Identity and Diversity" the *Essay* is introduced with the rubric '*Consciousness makes personal identity*' and it contains the sentence, "For it being the same consciousness that makes a man be himself to himself, personal identity depends on that only, whether it be annexed only to one individual substance, or can be continued in a succession of several substances." (See also §13 *passim*.) And so Locke has been interpreted by Reid and many other critics. But I have no wish to engage in a controversy about what Locke meant. Let the references to Locke in the present essay be read as references to "the Locke of the textbooks," a possibly historical, possibly fictional, but certainly important figure.

8 The following brief summary of Shoemaker's views is based on his well-known debate with Richard Swinburne on dualism and personal identity. See Sydney Shoemaker and Richard Swinburne, *Personal Identity* (Oxford: Basil Blackwell, 1984), pp. 108–10.

9 The argument that follows is a version of an argument I first presented in my essay "Materialism and the Psychological-Continuity Account of Personal Identity," *Philosophical Perspectives,* Vol. 11: *Mind, Causation, and World* (1997), pp. 305–19. This essay is reprinted in *Ontology, Identity and Modality: Essays in Metaphysics* (a collection of some of my essays on metaphysics), forthcoming from Cambridge University Press.

The argument for the incoherency of Locke's theory of personal identity set out earlier in the present essay is an adaptation of this argument.

10 Dean Zimmerman has tried to persuade me that Chisholm never actually *held* this view. Well, if he did not hold it, he at any rate (to borrow a phrase of Plantinga's) entertained it with a considerable degree of hospitality.

11 S. Shoemaker, "Self and Substance," *Philosophical Perspectives,* Vol. 11: *Mind, Causation, and World* (1997), pp. 283–304. See particularly pp. 300–1.

12 The "Privy Council" example is not Shoemaker's. It is suggested by his list of examples of things that are not autonomous self-perpetuators: "baseball teams, corporations, religious sects." I have used the Privy Council as an example because it might be argued that teams, corporations, and sects incorporate at least some immanent causation.

13 Or this is what would have to happen if, for any *x*s, those *x*s have at most one mereological sum at a time. If a given set of objects can simultaneously have more than one mereological sum, the following might be what happens. Before the transfer of information, both the 101-atoms and the 102-atoms have six mereological sums. One of them, the one that is Alice, is translated from one room to the other, and then the 101-atoms have five sums and the 102-atoms have seven sums.

14 For a discussion of the relation between bodily transfer and psychological continuity, see "Materialism and the Psychological-Continuity Account of Personal Identity," pp. 315–18.

15 See his "Survival and Identity," in David Lewis, *Philosophical Papers, Volume I* (New York: Oxford University Press, 1983), pp. 55–73. The paper was originally published in Amélie O. Rorty (ed.), *The Identities of Persons* (Berkeley: University of California Press, 1976).