

## Personal Identity: the Dualist Theory

From Sydney Shoemaker and Richard Swinburne, *Personal Identity* (Oxford: Blackwell, 1984)

### 1. Empiricist Theories

There are two philosophical questions about personal identity. The first is: what are the logically necessary and sufficient conditions for a person  $P_2$  at a time,  $t_2$  being the same person as a person  $P_1$  at an earlier time  $t_1$ ,<sup>1</sup> or, loosely, what does it mean to say that  $P_2$  is the same person as  $P_1$ ? The second is: what evidence of observation and experience can we have that a person  $P_2$  at  $t_2$  is the same person as a person  $P_1$  at  $t_1$  (and how are different pieces of evidence to be weighed against each other)? Many writers about personal identity have, however, needed to give only one account of personal identity, because their account of the logically necessary and sufficient conditions of personal identity was in terms of the evidence of observation and experience which would establish or oppose claims of personal identity. They have made no sharp distinction between the meaning of such claims and the evidence which supported them. Theories of this kind we may call empiricist theories.

In this section I shall briefly survey the empiricist theories which have been offered and argue that they are ultimately unsatisfactory, and so go on to argue that my two questions have very different answers. What we mean when we say that two persons are the

same is one thing; the evidence which we may have to support our claim is something very different.

The most natural theory of personal identity which readily occurs to people, is that personal identity is constituted by bodily identity.  $P_2$  is the same person as  $P_1$  if  $P_2$ 's body is the same body as  $P_1$ 's body. The person to whom you are talking now and call 'John' is the same person as the person to whom you were talking last week and then called 'John' if and only if he has the same body. To say that the two bodies – call them  $B_1$  and  $B_2$  – are the same is not to say that they contain exactly the same bits of matter. Bodies are continually taking in new matter (by people eating and drinking and breathing in) and getting rid of matter. But what makes the bodies the same is that the replacement of matter is only gradual. The matter which forms my body is organized in a certain way, into parts – legs, arms, heart, liver, etc. – which are interconnected and exchange matter and energy in regular ways. What makes my body today the same body as my body yesterday is that most of the matter is the same (although I may have lost some and gained some) and its organization has remained roughly the same.

This bodily theory of personal identity gives a somewhat similar account of personal identity to the account which it is natural to give of the identity of any material object or plant, and which is due ultimately to Aristotle (*Metaphysics*, Book 7). Aristotle distinguished between substances and properties. Substances are the individual things, like tables and chairs, cars and plants, which have properties (such as being square or round or red). Properties are 'universals', that is they can be possessed by many different substances; many different substances can be square or red. Substances are the individual substances which they are because of the matter out of which they are made and the form which is given to that matter. By 'the form' is meant those properties (normally of shape and organization) the possession of which is essential if a substance is to be the substance in question, the properties which it cannot lose without ceasing to exist. We thus distinguish between the essential properties of a substance – those which constitute its

---

<sup>1</sup> The logically necessary and sufficient conditions for something being so are those conditions such that if they are present, that thing must be so; and if they are absent, that thing cannot be so – all this because of considerations of logic.

form – and the accidental properties of a substance. It is among the essential properties of a certain oak tree that it has, under normal conditions, a certain general shape and appearance, a certain life cycle (of producing leaves in spring and acorns in autumn); but its exact height, its position, and the distribution of leaves on its tallest branch are accidental properties. If the matter of the oak tree is reduced to a heap of planks, the oak tree, lacking its essential properties, has ceased to exist. We think of substances as belonging to different kinds, natural – e.g., oak trees or ferns; or artificial – e.g., cars or desks; and the defining properties of a kind constitute the form of a substance which belongs to it. ...

What makes a substance the same substance as an earlier substance is that its matter is the same, or obtained from the matter of the former substance by gradual replacement, while continuing to possess the essential properties which constitute its form. The table at which I am writing today is the same table at which I was writing yesterday because it consists of the same matter (or at any rate, most of the same matter), organized in the same way – into the form of a table. For inanimate things, however, too much replacement of matter, however gradual, will destroy identity. If I replace the drawer of my desk by another drawer, the desk remains the same desk. But if, albeit gradually, I replace first the drawers and then the sides and then the top, so that there is none of the original matter left, we would say that the resulting desk was no longer the same desk as the original desk. For living things, such as plants, total replacement of matter – so long as it is gradual, and so long as physiology and anatomy also change only gradually if at all – will not destroy identity. The oak tree is the same as the sapling out of which it has grown, because replacement of matter has been gradual, and form (i.e., shape, physiology, and behaviour) has been largely preserved while any changes in it have been gradual. ...

Persons too are substances. (Men, or human beings, are persons of a certain kind – viz., those with similar anatomy, physiology, and evolutionary origin to ourselves. There may be persons, e.g., on

another planet, who are not human beings.) If we apply Aristotle's general account of the identity of substances to persons, it follows that for a person to be the same person as an earlier person, he has to have the same matter (or matter obtained from that earlier person by gradual replacement) organized into the form of a person. The essential properties which make the form of a person would include, for Aristotle, not merely shape and physiological properties, but a kind of way of behaving and a capacity for a mental life of thought and feeling. For  $P_2$  at  $t_2$  to be the same person as  $P_1$  at  $t_1$ , both have to be persons (to have a certain kind of body and mental life) and to be made of the same matter (i.e., to be such that  $P_2$ 's body is obtained from  $P_1$ 's by gradual replacement of parts). Such is the bodily theory of personal identity. It does not deny that persons have a mental life, but insists that what makes a person the same person as an earlier person is sameness of body.

The difficulty which has been felt by those modern philosophers basically sympathetic to a bodily theory of personal identity is this. One part of the body – viz. the brain – seems to be of crucial importance for determining the characteristic behaviour of the rest. The brain controls not merely the physiology of the body but the way people behave and talk and think. If a man loses an arm or a leg, we do not think that the subsequent person is in any way different from the original person. If a man has a heart transplant or a liver transplant, again we do not think that the replacement makes a different person. On the other hand, if the brain of a person  $P_1$  were removed from his body  $B_1$  and transplanted into the skull of a body  $B_2$  of a person  $P_2$ , from which the brain was removed and then transplanted into the empty skull of  $B_1$  (i.e., if brains were interchanged), we would have serious doubt whether  $P_1$  had any more the same body. We would be inclined to say that the person went where his brain went – viz., that  $P_1$  at first had body  $B_1$ , and then, after the transplant, body  $B_2$ . The reason why we would say this is that (we have very good scientific reason to believe) the person with  $B_2$ 's body would claim to be  $P_1$ , to have done and experienced the things which we know  $P_1$  to have done, and would have the character, beliefs, and attitudes of  $P_1$ . What

determines my attitude towards a person is not so much the matter out of which his body is made, but who he claims to be, whether he has knowledge of my past life purportedly on the basis of previous acquaintance with me, and more generally what his beliefs about the world are and what are his attitudes towards it. Hence a philosopher seeking a materialist criterion of personal identity, will come to regard the brain, the core of the body, rather than the rest of the body as what matters for personal identity. So this modified bodily theory states: that  $P_2$  is the same person as  $P_1$  if and only if  $P_2$  has the same central organ controlling memory and character, viz., same brain, as  $P_1$ . Let us call it the brain theory of personal identity. A theory along these lines (with a crucial qualification, to be discussed shortly) was tentatively suggested by David Wiggins in *Identity and Spatiotemporal Continuity* (Oxford, 1967).

The traditional alternative to a bodily theory of personal identity is the memory-character theory. This claims that, given the importance for our attitude towards persons of their memory claims and character, continuity in respect of these would constitute personal identity – whether or not this continuity is caused by continuity of some bodily organ, such as the brain; and the absence of continuity of memory and character in some particular case involves the absence of personal identity, even if there is continuity in respect of that bodily organ which produces such continuity between other persons on other occasions.

The simplest version of this theory was that given by John Locke. According to Locke, memory alone (or ‘consciousness’, as he often calls it) constitutes personal identity. Loosely –  $P_2$  at  $t_2$  is the same person as  $P_1$  at an earlier time  $t_1$ , if and only if  $P_2$  remembers having done and experienced various things, where these things were in fact done and experienced by  $P_1$ .

Before expounding Locke’s theory further we need to be clear about the kind of memory which is involved. First, it is what is sometimes called personal memory, i.e., memory of one’s own past experiences. It is thus to be distinguished from factual

memory, which is memory of some fact known previously; as when I remember that the battle of Hastings was fought in 1066. This is not a memory of a past experience. ... Secondly, it is personal memory in the weak sense. In the normal or strong sense of ‘remember’, one can only remember doing something if one really did it. I may say that I ‘remember’ going up the Eiffel Tower, but if I didn’t do it, it seems natural to say that I cannot really remember having done it. In this sense, just as you can only know what is true, so you can only remember what you really did. However, there is also a weak sense of ‘remember’ in which a man remembers whatever he believes that he remembers in the strong sense. One’s weak memories are not necessarily true ones. Now if the memory criterion defined personal identity in terms of memory in the strong sense, it would not be very useful; for to say that  $P_2$  remembers having done what  $P_1$  did would already entail their being the same person, and anyone in doubt as to whether  $P_2$  was the same person as  $P_1$ , would have equal doubt whether  $P_2$  really did remember doing what  $P_1$  did. What the criterion as stated is concerned with is memory in the weak sense, which (because the strong sense is the more natural one) I shall henceforward call apparent memory.

So Locke’s theory can now be rephrased as follows:  $P_2$  at  $t_2$  is the same person as  $P_1$  at an earlier time  $t_1$ , if and only if  $P_2$  apparently remembers having done and experienced various things when those things were in fact done and experienced by  $P_1$ . A person is who he thinks that he is. ...

Locke’s theory needs tidying up if we are to avoid absurdity. Consider, first, the following objection made by Thomas Reid:

Suppose a brave officer to have been flogged when a boy at school for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life; suppose also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that, when made a general, he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.

These things being supposed, it follows, from Mr Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging; therefore according to Mr Locke's doctrine, he is not the same person who was flogged. Therefore the general is, and at the same time is not, the same person with him who was flogged at school. (Reid, *Essays on the Intellectual Powers of Man*, bk. III, ch. 6)

The objection illustrates the important point that identity is a transitive relation; if  $a$  is identical with  $b$  and  $b$  is identical with  $c$ , then necessarily  $a$  is identical with  $c$ . We can meet the objection by reformulating Locke's theory as follows:  $P_2$  at  $t_2$  is the same person as  $P_1$  at an earlier time  $t_1$  if and only if *either*  $P_2$  apparently remembers what  $P_1$  did and experienced, or he apparently remembers what some person  $P'$  at an intermediate time  $t'$  did and experienced, when  $P'$  apparently remembers what  $P_1$  did and experienced, *or* they are linked by some longer intermediate chain. (That is,  $P_2$  apparently remembers what  $P'$  did and experienced,  $P'$  apparently remembers what  $P''$  did and experienced, and so on until we reach a person who apparently remembers what  $P_1$  did and experienced.) If  $P_1$  and  $P_2$  are linked by such a chain, they are, we may say, linked by continuity of memory. Clearly, the apparent memories of the deeds and experiences of the previous person at each stage in the chain need not be completely accurate memories of what was done and experienced. But they do need to be fairly accurate memories of what was done and experienced, if the later person is to be the person who did and experienced those things. ...

Many advocates of a memory theory have not always been very clear in their exposition about whether the apparent memories which form the links in the chain of memory need to be actual memories, or whether they need only to be hypothetical memories. By 'actual memories' I mean actual recallings of past experiences. The trouble with the suggestion that actual memories are required is that we do not very often recall our past, and it seems natural to

suppose that the deeds and experiences of some moments of a person's life never get re-called. Yet the memory theory, as stated so far, rules out that possibility. If I am not connected by a chain of memories with the deeds and experiences done by a person at a certain time, then I am not identical with that person. It is perhaps better if the theory claims that the apparent memories which form the links need only be hypothetical memories – i.e., what a person would apparently remember if he were to try to remember the deeds and experiences in question, e.g., in consequence of being prompted.

There is, however, a major objection to any memory theory of personal identity, arising from the possibility of duplication. The objection was made briefly by Reid and at greater length in an influential article by Bernard Williams. Williams imagines the case of a man whom he calls Charles who turns up in the twentieth-century claiming to be Guy Fawkes:

All the events he claims to have witnessed and all the actions he claims to have done point unanimously to the life-history of some one person in the past – for instance Guy Fawkes. Not only do all Charles' memory-claims that can be checked fit the pattern of Fawkes' life as known to historians, but others that cannot be checked are plausible, provide explanations of unexplained facts, and so on.<sup>2</sup>

The fact that memory claims which "cannot be checked are plausible, provide explanations of unexplained facts, and so on" is evidence that Charles is not merely claiming to remember what he has in fact read in a book about Guy Fawkes, and so leaves us back with the supposition, natural to make in normal cases, that he is reporting honestly his apparent memories. So, by a memory theory Charles would be Guy Fawkes. But then suppose, Williams imagines, that another man Robert turns up, who satisfies the memory criteria for being Guy Fawkes equally well. We cannot

---

<sup>2</sup> Bernard Williams, 'Personal Identity and Individuation', *Proceedings of the Aristotelian Society*, (1956-57), p. 332.

say that they are both identical with Guy Fawkes, for if they were, they would be identical with each other – which they are not since they currently live different lives and have different thoughts and feelings from each other. So apparent memory cannot constitute personal identity, although it may be fallible evidence of it.

The objection from the possibility of duplication, together with other difficulties which will be mentioned in later chapters, have inclined the majority of contemporary writers to favour a theory which makes some sort of bodily continuity central to personal identity. As we have seen, the brain theory takes into account the insight of memory-and-character theory into the importance of these factors for personal identity, by selecting the brain, as the organ causally responsible for the continuity of memory and character, as that part of the body the continuity of which constitutes the continuity of the person.

The trouble is that any brain theory is also open to the duplication objection. The human brain has two very similar hemispheres – a left and a right hemisphere. The left hemisphere plays a major role in the control of limbs of, and processing of sensory information from, the right side of the body (and from the right sides of the two eyes); and the right hemisphere plays a major role in the control of limbs of, and processing of sensory information from, the left side of the body (and from the left sides of the two eyes). The left hemisphere plays a major role in the control of speech. Although the hemispheres have different roles in the adult, they interact with each other; and if parts of a hemisphere are removed, at any rate early in life, the roles of those parts are often taken over by parts of the other hemisphere. Brain operations which remove substantial parts of the brain are not infrequent. It might be possible one day to remove a whole hemisphere, without killing the person. There are no logical difficulties in supposing that we could transplant one of  $P_1$ 's hemispheres into one skull from which a brain had been removed, and the other hemisphere into another such skull, and that both transplants should take, and it may well be practically possible to do so. It is certainly more likely to occur than the Guy

Fawkes story told by Williams! If these transplants took, clearly each of the resulting persons would behave to some extent like  $P_1$ , and indeed both would probably have some of the apparent memories of  $P_1$ . Each of the resulting persons would then be good candidates for being  $P_1$ .

After all, if one of  $P_1$ 's hemispheres had been destroyed and the other remained intact and untransplanted, and the resulting person continued to behave and make memory claims somewhat like those of  $P_1$ , we would have had little hesitation in declaring that person to be  $P_1$ . The same applies, whichever hemisphere was preserved – although it may well be that the resulting person would have greater capacities (e.g. speech) if one hemisphere was preserved than if the other one was preserved. We have seen earlier, good reason for supposing that the person goes where his brain goes, and if his brain consists only of one hemisphere, that should make no difference. So if the one remaining hemisphere is then transplanted, we ought to say that the person whose body it now controls is  $P_1$ . Whether that person is  $P_1$  can hardly be affected by the fact that instead of being destroyed, the other hemisphere is also transplanted so as to constitute the brain of person. But if it is, that other person will be just as good a candidate for being  $P_1$ . So a Wiggins-type account might lead us to say that both resulting persons are  $P_1$ . But, for the reason given earlier in connection with the Guy Fawkes examples, that cannot be – since the two later persons are not identical with each other. Hence, Wiggins adds to his tentative definition a clause stating that  $P_2$  who satisfies his criterion stated earlier is the same person as  $P_1$ , only if there is no other later person who also satisfies the criterion.

But the introduction into any theory, whether a memory theory, a brain theory, or whatever, of a clause stating that a person who satisfies the criterion in question for being the same as an earlier person is the same, only so long as there is no other person who satisfies the criterion also or equally well, does have an absurd consequence. Let us illustrate this for the brain theory. Suppose

P<sub>1</sub>'s left hemisphere is transplanted into some skull and the transplant takes. Then, according to the theory, whether the resulting person is P<sub>1</sub>, i.e., whether P<sub>1</sub> survives, will depend on whether the other transplant takes. If it does, since both resulting persons will satisfy the memory and brain continuity criteria equally well, neither will be P<sub>1</sub>. But if the other transplant does not take, then since there is only one person who satisfies the criterion, that person is P<sub>1</sub>. So whether I survive an operation will depend on what happens in a body entirely different from the body which will be mine, if I do survive. But how can who I am depend on what happens to you? A similar absurd consequence follows when a similar clause forbidding duplication is added to a memory theory.

Yet if we abandon the duplication clause, we are back with the original difficulty – that there may be more than one later person who satisfies any memory criterion or brain criterion, or combination thereof, for being the same person as an earlier person. Our discussion brings to our attention also the fact that both these criteria are criteria which may be satisfied to varying degrees. P<sub>2</sub> can have 90 per cent, or 80 per cent, or less than 50 per cent of the brain of P<sub>1</sub>; and likewise the similarity of apparent memory and character may vary along a spectrum. Just how well do criteria have to be satisfied for the later person to be the same person as the earlier person? Any line one might draw seems totally artificial. One might think that it was non-arbitrary to insist on more than 50 per cent of the original brain matter – for only one later person could have more than 50 per cent of the original brain matter (whereas if our criterion demands only a smaller proportion, more than one later person could satisfy it). But would we really want to say that P<sub>6</sub> was the same person as P<sub>1</sub> if P<sub>2</sub> was obtained from P<sub>1</sub> by a transplant of 60 per cent (and so more than half) of P<sub>1</sub>'s brain matter, P<sub>3</sub> was obtained from P<sub>2</sub> by a transplant of 60 per cent of P<sub>2</sub>'s brain matter, and so on until we came to P<sub>6</sub>. By the criterion of 'more than half of the brain matter', P<sub>6</sub> would be the same person as P<sub>5</sub>, P<sub>5</sub> as P<sub>4</sub> and so on, and so by the transitivity of identity P<sub>6</sub> would be the same person as P<sub>1</sub> – although he would have very little of P<sub>1</sub>'s brain matter. Any criterion of the proportion

of brain matter transferred, to be plausible, would have to take account of whether there had been similar transplants in the past, and the length of the interval between them. And then the arbitrariness of the criterion would stare us in the face.

This problem pushes the thinker towards one of two solutions. The first solution is to say that personal identity is a matter of degree. P<sub>2</sub> is the same person as P<sub>1</sub> to the extent to which there is sameness of brain matter and continuity of memory. After all, survival for inanimate things is a matter of degree. As we gradually replace bits of a desk with new bits, the resulting desk is only more or less the same as the original desk. And if my car is taken to pieces and some of the bits are used to make one new car, and some of the bits used to make another new car, both cars are partly the same as and partly different from the old car. Why cannot we say the same of people? Normally we are not inclined to talk thus, because brain operations are rare and brain hemisphere transplants never happen. Hence there is normally at most only one candidate for being the same person as an earlier person, and he is normally a very strong candidate indeed – having a more or less identical brain and very great similarities of apparent memory and character. So we tend to think of personal identity as all or nothing. But it is not thus in its logic, the argument goes. There is the logical possibility, which could become an empirical possibility, of intermediate cases of persons who are to some extent the same as and to some extent different from original persons.

This view has been advocated by Derek Parfit.<sup>3</sup> When a person divides, as a result of a split brain transplant, he 'survives' in part, Parfit holds, as each of two persons. They constitute his later 'selves', neither of whom, to speak strictly, are identical with the original person.

---

<sup>3</sup> "Personal Identity", *Philosophical Review*, 80 (1971), pp. 3-27.

This theory, which Parfit calls the complex view,<sup>4</sup> does, however, run up against a fundamental difficulty that it commits him to substantial empirical claims which to all appearance could very easily be false. I can bring this out by adopting Bernard Williams's famous mad surgeon story.<sup>5</sup> Suppose that a mad surgeon captures you and announces that he is going to transplant your left cerebral hemisphere into one body, and your right one into another. He is going to torture one of the resulting persons and free the other with a gift of a million pounds. You can choose which person is going to be tortured and which to be rewarded, and the surgeon promises to do as you choose. You believe his promise. But how are you to choose? You wish to choose that you are rewarded, but you do not know which resultant person will be you. Now on the complex theory each person will be you to the extent to which he has your brain and resembles you in his apparent memories and character. It would be in principle empirically ascertainable whether and to what extent persons with right hemisphere transplants resemble their originals in apparent memories and character more or less than persons with left hemisphere transplants. But clearly the difference is not going to be great. So Parfit must say that your choice does not greatly matter. Both subsequent persons will be in part you – although perhaps to slightly different degrees. And so you will – although perhaps to slightly different degrees – in part suffer and in part enjoy what each suffers and enjoys. So you have reason both for joyous expectation and for terrified anticipation. But one problem is: how could you have reason for part joyous expectation and part terrified anticipation, when no one future person is going to suffer a mixed fate?

---

<sup>4</sup> He introduces this terminology in his paper, 'On the Importance of Self-Identity', *Journal of Philosophy*, 68 (1971), pp. 683-90.

<sup>5</sup> Bernard Williams, 'The Self and the Future', *Philosophical Review*, 79 (1970), pp. 161-80.

But even if this notion of partial survival does make sense, the more serious difficulty remains, which is this. We can make sense of the supposition that the victim makes the wrong choice, and has the experience of being tortured and not the experience of being rewarded; or the right choice, and has the experience of being rewarded and not the experience of being tortured. A mere philosophical analysis of the concept of personal identity cannot tell you which experiences will be yours tomorrow. To use Bernard Williams's telling word, any choice would be a 'risk'. But on Parfit's view no risk would be involved – for knowing the extent of continuity of brain, apparent memory, and character, you would know the extent to which a future person would be you and so the extent to which his experiences would be yours. Although it *may* be the case that if my cerebral hemispheres are transplanted into different bodies, I survive partly as the person whose body is controlled by one and partly as the person whose body is controlled by the other, it may not be like that at all. Maybe I go where the left hemisphere goes; and when my right hemisphere is separated from the left hemisphere and comes to control a body by itself, either a new person is formed, or the resulting organism, although behaving to some extent like a person, is really a very complicated non-conscious machine. As we have noted, the fate of some parts of my body, such as my arms and legs, is quite irrelevant to the fate of me. And plausibly the fate of some parts of my brain is irrelevant – can I not survive completely a minor brain operation which removes a very small tumour? But then maybe it is the same with some larger parts of the brain too. We just don't know. If the mad surgeon's victim took the attitude that it didn't matter which way he chose, we must, I suggest, regard him as taking an unjustifiably dogmatic attitude.

The alternative way out of the duplication problem is to say that although apparent memory and brain continuity are, as they obviously are, evidence of personal identity, they are fallible evidence and personal identity is something distinct from them. Just as the presence of blood stains and fingerprints matching

those of a given man are evidence of his earlier presence at the scene of the crime, and the discovery of Roman-looking coins and buildings is evidence that the Romans lived in some region, so the similarity of  $P_2$ 's apparent memory to that of  $P_1$  and his having much the same brain matter, is evidence that  $P_2$  is the same person as  $P_1$ . Yet blood stains and fingerprints are one thing and a man's earlier presence at the scene of the crime another. His presence at the scene of the crime is not analysable in terms of the later presence of blood stains and fingerprints. The latter is evidence of the former, because you seldom get blood stains and fingerprints at a place, matching those of a given man, unless he has been there leaving them around. But it might happen. So, the suggestion is, personal identity is distinct from, although evidenced by, similarity of memory and continuity of brain.

This account, which for the moment I will follow Parfit in calling the simple view, can meet all the difficulties which have beset the other theories which we have discussed. The difficulty for the complex view was that it seemed very peculiar to suppose that mere logic could determine which of the experiences had by various persons, each of which was to some extent continuous with me in apparent memory and brain matter, would be mine. There seemed to be a further truth – that I would or would not have those experiences – beyond any truths about the extent of similarity in apparent memory and matter of future persons to myself. The simple view claims explicitly that personal identity is one thing, and the extent of similarity in matter and apparent memory another. There is no contradiction in supposing that the one should occur without the other. Strong similarity of matter and apparent memory is powerful evidence of personal identity. I and the person who had my body and brain last week have virtually the same brain matter and such similar apparent memory, that it is well-nigh certain that we are the same person. But where the brain matter is only in part the same and the memory connection less strong, it is only fairly probable that the persons are the same. Where there are two later persons  $P_2$  and  $P_2^*$ , each of whom had some continuity with the earlier person  $P_1$ , the evidence supports to some extent each of the

two hypotheses – that  $P_2$  is the same person as  $P_1$ , and that  $P_2^*$  is the same person as  $P_1$ . It may give more support to one hypothesis than to the other, but the less well supported hypothesis might be the true one, or maybe neither hypothesis is true. Perhaps  $P_1$  has ceased to exist, and two different persons have come into existence. So the simple view fully accepts that mere logic cannot determine which experiences will be mine, but it allows that continuity of apparent memory and brain provides fallible evidence about this. And of course the duplication objection that they allow for the two subsequent persons being the same person, which we brought against the brain and the memory theories, has no force against the simple theory. For although there can be equally good evidence that each of two later persons is the same person as an earlier person, that evidence is fallible; and since clearly only one person at one time can be strictly the same person as some person at an earlier time, it follows that in one case the evidence is misleading – although we may not know in which case.

. . . . In the next section I will expound and develop the simple view, and show that it amounts to the same as Cartesian dualism – the view that a person consists of two parts, soul, and body. . . .

## 2. The Dualist Theory

The brain transplant considerations of the first section leading to the simple view of personal identity showed that significant continuity of brain and memory was not enough to ensure personal identity. They did not show that continuity of brain or memory were totally dispensable; that  $P_2$  at time  $t_2$  could be the same person as  $P_1$  at an earlier time  $t_1$ , even though  $P_2$  had none of the brain matter (or other bodily matter) of  $P_1$  and had no apparent memory of  $P_1$ 's actions and experiences. A number of more extravagant thought-experiments do, however, show that there is no contradiction in this latter supposition.

There seems no contradiction in the supposition that a person might acquire a totally new body (including a completely new brain) – as many religious accounts of life after death claim that men do. To say that this body, sitting at the desk in my room is my body is to say two things. First it is to say that I can move parts of this body (arms, legs, etc.), just like that, without having to do any other intentional action and that I can make a difference to other physical objects only by moving parts of this body. By holding the door handle and turning my hand, I open the door. By bending my leg and stretching it I kick the ball and make it move into the goal. But I do not turn my hand or bend my leg by doing some other intentional action; I just do these things. Secondly, it is to say that my knowledge of states of the world outside this body is derived from their effects on this body – I learn about the positions of physical objects by seeing them, and seeing them involves light rays reflected by them impinging on my eyes and setting up nervous impulses in my optic nerve. My body is the vehicle of my agency in the world and my knowledge of the world. But then is it not coherent to suppose that I might suddenly find that my present body no longer served this function, that I could no longer acquire information through these eyes or move these limbs, but might discover that another body served the same function? I might find myself moving other limbs and acquiring information through other eyes. Then I would have a totally new body. If that body, like my last body, was an occupant of the Earth, then we would have a case of reincarnation, as Eastern religions have understood that. If that body was an occupant of some distant planet, or an environment which did not belong to the same space as our world, then we would have a case of resurrection as, on the whole, Western religions (Christianity, Judaism and Islam) have understood that. ...

Equally coherent, I suggest, is the supposition that a person might become disembodied. A person has a body if there is one particular chunk of matter through which he has to operate on and learn about the world. But suppose that he finds himself able to operate on and learn about the world within some small finite

region, without having to use one particular chunk of matter for this purpose. He might find himself with knowledge of the position of objects in a room (perhaps by having visual sensations, perhaps not), and able to move such objects just like that, in the ways in which we know about the positions of our limbs and can move them. But the room would not be, as it were, the person's body; for we may suppose that simply by choosing to do so he can gradually shift the focus of his knowledge and control, e.g., to the next room. The person would be in no way limited to operating and learning through one particular chunk of matter. Hence we may term him disembodied. The supposition that a person might become disembodied also seems coherent.

I have been arguing so far that it is coherent to suppose that a person could continue to exist with an entirely new body or with no body at all. ... Could a person continue to exist without any apparent memory of his previous doings? Quite clearly, we do allow not merely the logical possibility, but the frequent actuality of amnesia – a person forgetting all or certain stretches of his past life. Despite Locke, many a person does forget much of what he has done. But, of course, we normally only suppose this to happen in cases where there is the normal bodily and brain continuity. Our grounds for supposing that a person forgets what he has done are that the evidence of bodily and brain continuity suggests that he was the previous person who did certain things which he now cannot remember having done. And in the absence of both of the main kinds of evidence for personal identity, we would not be justified in supposing that personal identity held. ... For that reason I cannot describe a case where we would have good reason to suppose that  $P_2$  was identical with  $P_1$  even though there was neither brain continuity nor memory continuity between them. However, only given verificationist dogma is there any reason to suppose that the only things which are true are those of whose truth we can have evidence, and I shall suggest in section 3 [not included here] that there is no good reason for believing verificationism to be true. We can make sense of states of affairs being true, of which we can have no evidence that they are true.

And among them surely is the supposition that the person who acquires another body loses not merely control of the old one, but memories of what he did with its aid. ...

Those who hope to survive their death, despite the destruction of their body, will not necessarily be disturbed if they come to believe that they will then have no memory of their past life on Earth; they may just want to survive and have no interest in continuing to recall life on Earth. Again, apparently, there seems to be no contradiction involved in their belief. It seems to be a coherent belief (whether or not true or justified). Admittedly, there may be stories or beliefs which involve a hidden contradiction when initially they do not seem to do so. But the fact that there seems (and to so many people) to be no contradiction hidden in these stories is good reason for supposing that there is no contradiction hidden in them – until a contradiction is revealed. If this were not a good reason for believing there to be no contradiction, we would have no good reason for believing any sentence at all to be free of hidden contradiction. ...

In section 1, I set out Aristotle's account of the identity of substances: that a substance at one time is the same substance as a substance at an earlier time if and only if the later substance has the same form as, and continuity of matter with, the earlier substance. On this view a person is the same person as an earlier person if he has the same form as the earlier person (i.e., both are persons) and has continuity of matter with him (i.e., has the same body).

Certainly, to be the same person as an earlier person, a later person has to have the same form – i.e., has to be a person. If my arguments for the logical possibility of there being disembodied persons are correct, then the essential characteristics of a person constitute a narrower set than those which Aristotle would have included. My arguments suggest that all that a person needs to be a person are certain mental capacities – for having conscious experiences (i.e., thoughts or sensations) and performing

intentional actions. Thought-experiments of the kind described earlier allow that a person might lose his body, but they describe his continuing to have conscious experiences and his performing or being able to perform intentional actions, i.e., to do actions which he means to do, bring about effects for some purpose.

Yet if my arguments are correct, showing that two persons can be the same, even if there is no continuity between their bodily matter, we must say that in the form stated the Aristotelian account of identity applies only to inanimate objects and plants and has no application to personal identity.<sup>8</sup> We are then faced with a choice either of saying that the criteria of personal identity are different from those for other substances, or of trying to give a more general account than Aristotle's of the identity of substances which would cover both persons and other substances. It is possible to widen the Aristotelian account so that we can do the latter. We have only to say that two substances are the same if and only if they have the same form and there is continuity of the stuff of which they are made, and allow that there may be kinds of stuff other than matter. I will call this account of substance identity the wider Aristotelian account. We may say that there is a stuff of another kind, immaterial stuff, and that persons are made of both normal bodily matter and of this immaterial stuff but that it is the continuity of the latter which provides that continuity of stuff which is necessary for the identity of the person over time.

This is in essence the way of expressing the simple theory which is adopted by those who say that a person living on Earth consists of two parts – a material part, the body; and an immaterial part, the soul. The soul is the essential part of a person, and it is its continuing which constitutes the continuing of the person. While on Earth, the soul is linked to a body (by the body being the vehicle of the person's knowledge of and action upon the physical world). But, it is logically possible, the soul can be separated from the body and exist in a disembodied state (in the way described earlier) or linked to a new body. This way of expressing things has been used in many religious traditions down the centuries, for it is

a very natural way of expressing what is involved in being a person once you allow that a person can survive the death of his body. Classical philosophical statements of it are to be found in Plato and, above all, in Descartes. I shall call this view classical dualism.

I wrote that ‘in essence’ classical dualism is the view that there is more stuff to the person than bodily matter, and that it is the continuing of this stuff which is necessary for the continuing of the person, because a writer such as Descartes did not distinguish between the immaterial stuff, let us call it soul-stuff, and that stuff being organized (with or without a body) as one soul. Descartes and other classical dualists however did not make this distinction, because they assumed (implicitly) that it was not logically possible that persons divide – i.e., that an earlier person could be in part the same person as each of two later persons. Hence they implicitly assumed that soul-stuff comes in essentially indivisible units. That is indeed what one has to say about soul-stuff, if one makes the supposition (as I was inclined to do, in section I) that it is not logically possible that persons divide. There is nothing odd about supposing that soul-stuff comes in essentially indivisible units. Of any chunk of matter, however small, it is always logically, if not physically, possible that it be divided into two. Yet it is because matter is extended, that one can always make sense of it being divided. For a chunk of matter necessarily takes up a finite volume of space. A finite volume of space necessarily is composed of two half-volumes. So it always makes sense to suppose that part of the chunk which occupies the left half-volume of space to be separated from that part of the chunk which occupies the right half-volume. But that kind of consideration has no application to immaterial stuff. There is no reason why there should not be a kind of immaterial stuff which necessarily is indivisible; and if the supposition of section I is correct, the soul-stuff will have that property. ...

Given that for any present person who is currently conscious, there is no logical impossibility, whatever else may be true now of that person, that that person continue to exist without his body, it

follows that that person must now actually have a part other than a bodily part which can continue, and which we may call his soul – and so that his possession of it is entailed by his being a conscious thing. For there is not even a logical possibility, that if I now consist of nothing but matter and the matter is destroyed, I should nevertheless continue to exist. From the mere logical possibility of my continued existence there follows the actual fact that there is now more to me than my body; and that more is the essential part of myself. A person’s being conscious is thus to be analysed as an immaterial core of himself, his soul being conscious.

So Descartes argues, and his argument seems to me correct – given the wider Aristotelian framework. If we are prepared to say that substances can be the same, even though none of the stuff (in a wide sense) of which they are made is the same, the conclusion does not follow. The wider Aristotelian framework provides a partial definition of ‘stuff’ rather than a factual truth.

To say that a person has an immaterial soul is not to say that if you examine him closely enough under an acute enough microscope you will find some very rarefied constituent which has eluded the power of ordinary microscopes. It is just a way of expressing the point within a traditional framework of thought that persons can – it is logically possible – continue, when their bodies do not. It does, however, seem a very natural way of expressing the point – especially once we allow that persons can become disembodied. Unless we adopt a wider Aristotelian framework, we shall have to say that there can be substances which are not made of anything, and which are the same substances as other substances which are made of matter.

It does not follow from all this that a person’s body is no part of him. Given that what we are trying to do is to elucidate the nature of those entities which we normally call ‘persons’, we must say that arms and legs and all other parts of the living body are parts of the person. My arms and legs are parts of me. ...