1

Personal Identity

Introduction

It seems to be an obvious truth that we are directly morally responsible only for actions that *we* perform ourselves. If Eric robs a bank single-handed, then it is Eric who is responsible for the robbery and nobody else. If Eric has an identical twin brother, Ernie, who is physically indistinguishable from him and who has similar beliefs, mannerisms, personality and so on, it is still Eric who is responsible – no matter how similar to him Ernie is.

But what *makes* it the case that, after the robbery, one of the twins and not the other is the *same person as* the person who committed the crime? One possible answer is that sameness of person goes along with sameness of body. While Eric and Ernie are physically very similar, their bodies are numerically distinct: Eric's and Ernie's bodies are different bodies. If we had somehow inserted a tracking device under Eric's skin prior to the robbery, we would later know which of Eric and Ernie had committed the crime by discovering which one of them housed the device (assuming, of course, that Eric hasn't discovered it and surreptitiously transferred it to Ernie).

One might wonder, however, whether sameness of body – bodily identity – really is either necessary or sufficient for *personal* identity. As it happens, people cannot swap bodies with one another. But one might regard this as merely a contingent truth, so that while *in fact* judgements of personal identity and judgements of bodily identity go together, it doesn't follow that personal identity *consists* in bodily identity. Imagine that it is somehow possible to 'swap bodies', in a way that leaves all one's psychological features intact, by transplanting, say, *A*'s brain into *B*'s body and *B*'s brain into *A*'s body. (Of course, the brain is part of the body. But think of the 'body', for current purposes, as the body minus the brain. Or, alternatively, you might imagine that we could 'rewire' *A*'s and *B*'s brains, so that no brain transplant is necessary.)

Before the operation, then, we have two persons, *A* and *B*, and two bodies: call these 'the *A*-body' and 'the *B*-body' respectively. Imagine that *A* had an unhappy childhood and now works as an English teacher. She regrets never having been to the Bahamas and intends to go there next year. *B* had a happy childhood, has a boring job in a supermarket, and (to spice up her dull life) recently robbed a bank and took an extended holiday in the Bahamas on the proceeds – but she feels rather guilty about the robbery.

Now imagine that, after waking up, the person now 'inhabiting' the *A*-body – call her 'the *A*-body person' – is psychologically exactly like *B* was before the operation. She has all the same beliefs and personality traits as *B* used to have. Moreover, she can vividly remember everything about *B*'s past: events in *B*'s happy childhood, her boring job, the excitement of robbing the bank, and the holiday in the Bahamas. She feels remorse for the robbery, just as *B* did. Similarly, the person now inhabiting the *B*-body ('the *B*-body person') is psychologically just like *A* used to be. She remembers *A*'s unhappy childhood, is well-acquainted with the works of Shakespeare, regrets never having been to the Bahamas, and intends to go next year, just as *A* did before the operation.

Now, which of these two persons – the *A*-body person and the *B*-body person – *is B*? Or, to make the question more vivid, which of them do you regard as morally responsible for the robbery? Most people would say that it is the *A*-body person. After all, it is she who remembers the robbery and feels guilty about it. The *B*-body person, by contrast, has no memory whatever of robbing any bank, and if asked whether she did it, she would sincerely deny it.

If it is right to say that the *A*-body person is *B* – the bank-robber – then personal identity cannot consist in bodily identity, since a 'bodily criterion' of personal identity delivers the result that it is the *B*-body person who is *B*. So perhaps what makes someone the same person over time is whatever it is that leads us to judge that the *A*-body person is *B* – and that seems to be a matter of psychological rather than physical criteria.

As an introduction to the two texts reproduced here, we shall describe three problems that a theory of personal identity based on psychological rather than physical criteria has to solve. The first two are very old problems, first raised against John Locke's account of personal identity (1690). Locke's theory is often characterized as follows (though it is debatable whether this is exactly what he had in mind):

(M) *X* (at a later time) is the same person as *Y* (at an earlier time) if and only if *X* can remember *Y*'s experiences.

Because it appeals just to memory, (M) is often known as the 'memory theory' of personal identity.

Problem 1: The transitivity of identity

The first problem faced by the memory theory was originally raised by Thomas Reid (1785). Suppose a young boy (*A*), at time t_1, steals apples from an orchard. Some years later, at time t_2, an officer (*B*) remembers stealing the apples. So, by (M), *B* is the same

person as *A*. Many years after that, a general (*C*) can remember the officer's experiences. So, by (M), *C* is the same person as *B*. But *C* cannot remember stealing the apples. So, again by (M), *C* is not the same person as *A*.

Now, identity is what is known as a *transitive relation*. A transitive relation is a relation *R* such that if *X* bears *R* to *Y* and *Y* bears *R* to *Z*, it follows that *X* bears *R* to *Z*. So, for example, *is an ancestor of* is a transitive relation: if George is an ancestor of Mary, and Mary is an ancestor of John, it follows that George is an ancestor of John. Similarly for identity: if Bruce Wayne is Batman, and Batman is the Caped Crusader, it follows that Bruce Wayne is the Caped Crusader. *Is a cousin of*, by contrast, is not transitive: if Sue is a cousin of Sam and Sam is a cousin of Dr. Jones, it doesn't follow that Sue is a cousin of Dr Jones. (Sue might be Dr. Jones's sister, or Sue might *be* Dr. Jones.)

The problem for the memory theory is that *remembers the experiences of* is not a transitive relation: in the example given above, *C* remembers *B*'s experiences and *B* remembers *A*'s experiences, but *C* does not remember *A*'s experiences. So personal identity cannot *be* the *remembers the experiences of* relation. (M) delivers the result that $A = B$ and $B = C$, but $A \neq C$, and this cannot be true, because it violates the transitivity requirement: if $A = B$ and $B = C$, then *A must* be identical with *C*. The challenge for someone who wants to hold that personal identity consists in the holding of psychological relations is thus to come up with *transitive* psychological relations which might reasonably be thought to be what personal identity consists in. Parfit addresses this issue in the first text reproduced below.

Problem 2: The circularity objection

This problem, first raised against Locke by Joseph Butler (1736), is that it seems to be a *precondition* of *X*'s remembering *Y*'s actions that *X* and *Y* are the same person. It is a conceptual truth – a truth about the concept of memory – that if *X* is not the same person as *Y*, then *X* cannot genuinely remember *Y*'s experiences. If *X* and *Y* are different people, then, while *X* can have memory-*like* experiences and mistakenly *think* that they are memories of *Y*'s experiences – a madman can mistakenly think that he can remember leading the French troops at the battle of Waterloo – those memory-like experiences cannot be genuine memories. Only Napoleon could genuinely *remember* leading the French troops into battle.

Now, the point of this objection is not that it renders (M) *false*; it doesn't. Rather, the problem is that according to the memory theory, memory is supposed to provide a *criterion* or *definition* or *analysis* of personal identity: it is supposed to *explain* what makes it the case that *X* is the same person as *Y*. But it cannot do this if facts about memory themselves depend upon facts about personal identity; and the circularity objection allegedly shows that facts about memory *do* depend on facts about personal identity. If someone sincerely claims to remember winning the 2004 Best Director Oscar for *The Return of the King*, we cannot infer with certainty that he is Peter Jackson; he might be someone completely different and just be suffering from a delusion. In order to be able to infer with certainty that he really is Peter Jackson, we have to ascertain that he really does *remember* winning the Oscar. But we can only do

that by *first* ascertaining that he is Peter Jackson. So whether or not the person remembers winning the Oscar cannot *determine* whether or not he is Peter Jackson. In the first text reproduced below, Parfit attempts (*inter alia*) to meet the circularity objection; in the second text, Schechtman argues that Parfit's attempt fails.

Problem 3: Divided brains

The recent literature on personal identity is full of outlandish thought-experiments. We are asked to imagine a very odd science-fiction scenario, and then asked whether our intuitive judgements about personal identity in that scenario are the same as those delivered by a particular theory of personal identity. One such thought-experiment is what we'll call 'the divided-brain case'. Imagine that *A*'s brain is cut in half and each half is transplanted into a new body: one half into the *B*-body and the other into the *C*-body. (*A*'s body is destroyed.) Imagine that having half of *A*'s brain results in no psychological loss, so that *each* of the people who wake up after the operation (*B* and *C*) is psychologically continuous with *A*: each of them can remember *A*'s childhood, has the same hopes and fears and intentions as *A* did, feels remorse about some of *A*'s past actions, and so on.

If we hold that psychological continuity is sufficient for personal identity, it seems that we have to conclude that *B* is the same person as *A*, but also that *C* is the same person as *A*. But this is surely impossible. *B* and *C* are, it seems, numerically distinct persons. Ten years later, *B* might rob a bank; and, intuitively, *C* would not bear any moral responsibility for the robbery. But in that case, it cannot be the case both that $B = A$ *and* that $C = A$, because they together entail that $B = C$. (This is just the transitivity of identity again: if $B = A$ and $A = C$ then $B = C$.) Identity is, as Parfit puts it, 'logically one–one': one person cannot be identical to two distinct people.

One way to respond to the divided-brain case would be to give up on psychological continuity altogether, and hold that personal identity must after all be analysed in terms of bodily identity. (One might be especially inclined to draw this conclusion if one thinks that at least one of problem 1 and problem 2 cannot be solved satisfactorily.) However, Parfit argues instead that it is not personal *identity* that matters after all. What matters – for example, for the purposes of allocating moral responsibility – is the fact that *B* is psychologically continuous with *A*, and also that *C* is psychologically continuous with *A*. For example, if *A* committed a terrible crime just before undergoing the operation, we should hold *both B* and *C* responsible for it, and it would be appropriate for both of them to feel remorse. But what matters cannot therefore be *identity*, because there is no good answer to the question, 'which of *B* and *C* is the same person as *A*?' Parfit calls what matters 'survival': *A* 'survives' both as *B* and as *C*, but is identical with neither of them.

The texts

Parfit's overall aim in the ground-breaking paper 'Personal Identity' (1973), reproduced below, is to provide an account of 'survival', in terms of what he calls

'psychological connectedness', and to argue that we should think of our own lives in terms of survival rather than identity. Psychological connectedness comes in degrees: you can be more or less psychologically connected to a 'past self', depending on the extent to which you can remember the experiences of that past self, you still intend to realize your past self's intentions, you hold the same beliefs and desires, and so on. So, for example, rather than the general in problem 1 above thinking of himself as *identical* to both the apple-stealing boy and the officer, he should think of one of his past selves – the boy – as surviving as his present self to a very small degree (or perhaps not at all), while another, more recent past self – the captain – has survived to a much greater degree.

Because Parfit's account of survival appeals to psychological relations, including the *remembers-the-experiences-of* relation, he needs to find a way of meeting the circularity problem (problem 2 above), which is just as much a problem for Parfit's view as it is for the memory theory of personal identity. This he attempts to do; however, in her paper 'Personhood and Personal Identity' (1990), Marya Schechtman argues that Parfit's attempt fails. She concludes that we should abandon the search for a non-circular psychological criterion of personal identity (and, by extension, survival) altogether, but attempts to diagnose why we are inclined to find the mistaken thought that there is such a criterion so seductive.

Derek Parfit, 'Personal Identity'[1]

We can, I think, describe cases in which, though we know the answer to every other question, we have no idea how to answer a question about personal identity. These cases are not covered by the criteria of personal identity that we actually use.

Do they present a problem?

It might be thought that they do not, because they could never occur. I suspect that some of them could. (Some, for instance, might become scientifically possible.) But I shall claim that even if they did they would present no problem.

My targets are two beliefs: one about the nature of personal identity, the other about its importance.

The first is that in these cases the question about identity must have an answer.

a → No one thinks this about, say, nations or machines. Our criteria for the identity of these do not cover certain cases. No one thinks that in these cases the questions ''Is it the same nation?'' or ''Is it the same machine?'' must have answers.

[1] I have been helped in writing this by D. Wiggins, D. F. Pears, P. F. Strawson, A. J. Ayer, M. Woods, N. Newman, and (through his publications) S. Shoemaker.

Some people believe that in this respect they are different. They agree that our criteria of personal identity do not cover certain cases, but they believe that the nature of their own identity through time is, somehow, such as to guarantee that in these cases questions about their identity must have answers. This belief might be expressed as follows: "Whatever happens between now and any future time, either I shall still exist, or I shall not. Any future experience will either be *my* experience, or it will not."

This first belief – in the special nature of personal identity – has, I think, certain effects. It makes people assume that the principle of self-interest is more rationally compelling than any moral principle. And it makes them more depressed by the thought of aging and of death.

I cannot see how to disprove this first belief. I shall describe a problem case. But this can only make it seem implausible.

Another approach might be this. We might suggest that one cause of the belief is the projection of our emotions. When we imagine ourselves in a problem case, we do feel that the question "Would it be me?" must have an answer. But what we take to be a bafflement about a further fact may be only the bafflement of our concern.

I shall not pursue this suggestion here. But one cause of our concern is the belief which is my second target. This is that unless the question about identity has an answer, we cannot answer certain important questions (questions about such matters as survival, memory, and responsibility).

Against this second belief my claim will be this. Certain important questions do presuppose a question about personal identity. But they can be freed of this presupposition. And when they are, the question about identity has no importance.

I

We can start by considering the much-discussed case of the man who, like an amoeba, divides.[2]

b→ Wiggins has recently dramatized this case.[3] He first referred to the operation imagined by Shoemaker.[4] We suppose that my brain is transplanted into someone else's (brainless) body, and that the resulting person

[2] Implicit in John Locke, *Essay Concerning Human Understanding*, ed. by John W. Yolton (London, 1961), vol. II, Ch. XXVII, sec. 18, and discussed by (among others) A. N. Prior in "Opposite Number," *Review of Metaphysics*, 11 (1957–1958), and "Time, Existence and Identity," *Proceedings of the Aristotelian Society*, LVII (1965–1966); J. Bennett in "The Simplicity of the Soul," *Journal of Philosophy*, LXIV (1967); and R. Chisholm and S. Shoemaker in "The Loose and Popular and the Strict and the Philosophical Senses of Identity," in *Perception and Personal Identity: Proceedings of the 1967 Oberlin Colloquium in Philosophy*, ed. by Norman Care and Robert H. Grimm (Cleveland, 1967).

[3] In *Identity and Spatio-Temporal Continuity* (Oxford, 1967), p. 50.

[4] In *Self-Knowledge and Self-Identity* (Ithaca, NY, 1963), p. 22.

has my character and apparent memories of my life. Most of us would agree, after thought, that the resulting person is me. I shall here assume such agreement.[5]

Wiggins then imagined his own operation. My brain is divided, and each half is housed in a new body. Both resulting people have my character and apparent memories of my life.

What happens to me? There seem only three possibilities: (1) I do not survive; (2) I survive as one of the two people; (3) I survive as both.

The trouble with (1) is this. We agreed that I could survive if my brain were successfully transplanted. And people have in fact survived with half their brains destroyed. It seems to follow that I could survive if half my brain were successfully transplanted and the other half were destroyed. But if this is so, how could I *not* survive if the other half were also successfully transplanted? How could a double success be a failure?

We can move to the second description. Perhaps one success is the maximum score. Perhaps I shall be one of the resulting people.

The trouble here is that in Wiggins' case each half of my brain is exactly similar, and so, to start with, is each resulting person. So how can I survive as only one of the two people? What can make me one of them rather than the other?

It seems clear that both of these descriptions – that I do not survive, and that I survive as one of the people – are highly implausible. Those who have accepted them must have assumed that they were the only possible descriptions.

What about our third description: that I survive as both people?

It might be said, ''If 'survive' implies identity, this description makes no sense – you cannot be two people. If it does not, the description is irrelevant to a problem about identity.''

I shall later deny the second of these remarks. But there are ways of denying the first. We might say, ''What we have called 'the two resulting people' are not two people. They are one person. I do survive Wiggins' operation. Its effect is to give me two bodies and a divided mind.''

It would shorten my argument if this were absurd. But I do not think it is. It is worth showing why.

We can, I suggest, imagine a divided mind. We can imagine a man having two simultaneous experiences, in having each of which he is unaware of having the other.

We may not even need to imagine this. Certain actual cases, to which Wiggins referred, seem to be best described in these terms. These involve

[5] Those who would disagree are not making a mistake. For them my argument would need a different case. There must be some multiple transplant, faced with which these people would both find it hard to believe that there must be an answer to the question about personal identity, and be able to be shown that nothing of importance turns upon this question.

the cutting of the bridge between the hemispheres of the brain. The aim was to cure epilepsy. But the result appears to be, in the surgeon's words, the creation of ''two separate spheres of consciousness,''[6] each of which controls one half of the patient's body. What is experienced in each is, presumably, experienced by the patient.

There are certain complications in these actual cases. So let us imagine a simpler case.

C → Suppose that the bridge between my hemispheres is brought under my voluntary control. This would enable me to disconnect my hemispheres as easily as if I were blinking. By doing this I would divide my mind. And we can suppose that when my mind is divided I can, in each half, bring about reunion.

This ability would have obvious uses. To give an example: I am near the end of a maths exam, and see two ways of tackling the last problem. I decide to divide my mind, to work, with each half, at one of two calculations, and then to reunite my mind and write a fair copy of the best result.

What shall I experience?

When I disconnect my hemispheres, my consciousness divides into two streams. But this division is not something that I experience. Each of my two streams of consciousness seems to have been straightforwardly continuous with my one stream of consciousness up to the moment of division. The only changes in each stream are the disappearance of half my visual field and the loss of sensation in, and control over, half my body.

Consider my experiences in what we can call my ''right-handed'' stream. I remember that I assigned my right hand to the longer calculation. This I now begin. In working at this calculation I can see, from the movements of my left hand, that I am also working at the other. But I am not aware of working at the other. So I might, in my right-handed stream, wonder how, in my left-handed stream, I am getting on.

My work is now over. I am about to reunite my mind. What should I, in each stream, expect? Simply that I shall suddenly seem to remember just having thought out two calculations, in thinking out each of which I was not aware of thinking out the other. This, I submit, we can imagine. And if my mind was divided, these memories are correct.

In describing this episode, I assumed that there were two series of thoughts, and that they were both mine. If my two hands visibly wrote out two calculations, and if I claimed to remember two corresponding series of thoughts, this is surely what we should want to say.

If it is, then a person's mental history need not be like a canal, with only one channel. It could be like a river, with islands, and with separate streams.

To apply this to Wiggins' operation: we mentioned the view that it gives me two bodies and a divided mind. We cannot now call this absurd. But it is, I think, unsatisfactory.

[6] R. W. Sperry, in *Brain and Conscious Experience*, ed. by J. C. Eccles (New York, 1966), p. 299.

There were two features of the case of the exam that made us want to say that only one person was involved. The mind was soon reunited, and there was only one body. If a mind was permanently divided and its halves developed in different ways, the point of speaking of one person would start to disappear. Wiggins' case, where there are also two bodies, seems to be over the borderline. After I have had his operation, the two ''products'' each have all the attributes of a person. They could live at opposite ends of the earth. (If they later met, they might even fail to recognize each other.) It would become intolerable to deny that they were different people.

Suppose we admit that they are different people. Could we still claim that I survived as both, using ''survive'' to imply identity?

We could. For we might suggest that two people could compose a third. We might say, ''I do survive Wiggins' operation as two people. They can be different people, and yet be me, in just the way in which the Pope's three crowns are one crown.''[7]

This is a possible way of giving sense to the claim that I survive as two different people, using ''survive'' to imply identity. But it keeps the language of identity only by changing the concept of a person. And there are obvious objections to this change.[8]

The alternative, for which I shall argue, is to give up the language of identity. We can suggest that I survive as two different people without implying that I am these people.

When I first mentioned this alternative, I mentioned this objection: ''If your new way of talking does not imply identity, it cannot solve our problem. For that is about identity. The problem is that all the possible answers to the question about identity are highly implausible.''

We can now answer this objection.

We can start by reminding ourselves that this is an objection only if we have one or both of the beliefs which I mentioned at the start of this paper.

The first was the belief that to any question about personal identity, in any describable case, there must be a true answer. For those with this belief, Wiggins' case is doubly perplexing. If all the possible answers are implausible, it is hard to decide which of them is true, and hard even to keep the belief that one of them must be true. If we give up this belief, as I think we should, these problems disappear. We shall then regard the case as like many others in which, for quite unpuzzling reasons, there *is* no answer

[7] Cf. Wiggins, p. 40.

[8] Suppose the resulting people fight a duel. Are there three people fighting, one on each side, and one on both? And suppose one of the bullets kills. Are there two acts, one murder and one suicide? How many people are left alive? One? Two? (We could hardly say, ''One and a half.'') We could talk in this way. But instead of saying that the resulting people *are* the original person – so that the pair is a trio – it would be far simpler to treat them as a pair, and describe their relation to the original person in some new way. (I owe this suggested way of talking, and the objections to it, to Michael Woods.)

to a question about identity. (Consider "Was England the same nation after 1066?")

Wiggins' case makes the first belief implausible. It also makes it trivial. For it undermines the second belief. This was the belief that important questions turn upon the question about identity. (It is worth pointing out that those who have only this second belief do not think that there must *be* an answer to this question, but rather that we must decide upon an answer.)

Against this second belief my claim is this. Certain questions do presuppose a question about personal identity. And because these questions *are* important, Wiggins' case does present a problem. But we cannot solve this problem by answering the question about identity. We can solve this problem only by taking these important questions and prizing them apart from the question about identity. After we have done this, the question about identity (though we might for the sake of neatness decide it) has no further interest.

Because there are several questions which presuppose identity, this claim will take some time to fill out.

d→ We can first return to the question of survival. This is a special case, for survival does not so much presuppose the retaining of identity as seem equivalent to it. It is thus the general relation which we need to prize apart from identity. We can then consider particular relations, such as those involved in memory and intention.

"Will I survive?" seems, I said, equivalent to "Will there be some person alive who is the same person as me?"

If we treat these questions as equivalent, then the least unsatisfactory description of Wiggins' case is, I think, that I survive with two bodies and a divided mind.

Several writers have chosen to say that I am neither of the resulting people. Given our equivalence, this implies that I do not survive, and hence, presumably, that even if Wiggins' operation is not literally death, I ought, since I will not survive it, to regard it *as* death. But this seemed absurd.

It is worth repeating why. An emotion or attitude can be criticized for resting on a false belief, or for being inconsistent. A man who regarded Wiggins' operation as death must, I suggest, be open to one of these criticisms.

He might believe that his relation to each of the resulting people fails to contain some element which is contained in survival. But how can this be true? We agreed that he *would* survive if he stood in this very same relation to only *one* of the resulting people. So it cannot be the nature of this relation which makes it fail, in Wiggins' case, to be survival. It can only be its duplication.

Suppose that our man accepts this, but still regards division as death. His reaction would now seem wildly inconsistent. He would be like a man who, when told of a drug that could double his years of life, regarded the taking of this drug as death. The only difference in the case of division is that the extra years are to run concurrently. This is an interesting difference. But it cannot mean that there are *no* years to run.

I have argued this for those who think that there must, in Wiggins' case, be a true answer to the question about identity. For them, we might add, ''Perhaps the original person does lose his identity. But there may be other ways to do this than to die. One other way might be to multiply. To regard these as the same is to confuse nought with two.''

For those who think that the question of identity is up for decision, it would be clearly absurd to regard Wiggins' operation as death. These people would have to think, ''We could have chosen to say that I should be one of the resulting people. If we had, I should not have regarded it as death. But since we have chosen to say that I am neither person, I *do*.'' This is hard even to understand.[9]

My first conclusion, then, is this. The relation of the original person to each of the resulting people contains all that interests us – all that matters – in any ordinary case of survival. This is why we need a sense in which one person can survive as two.[10]

One of my aims in the rest of this paper will be to suggest such a sense. But we can first make some general remarks.

II

Identity is a one–one relation. Wiggins' case serves to show that what matters in survival need not be one–one.

e→ Wiggins' case is of course unlikely to occur. The relations which matter are, in fact, one–one. It is because they are that we can imply the holding of these relations by using the language of identity.

This use of language is convenient. But it can lead us astray. We may assume that what matters *is* identity and, hence, has the properties of identity.

In the case of the property of being one–one, this mistake is not serious. For what matters is in fact one–one. But in the case of another property, the mistake *is* serious. Identity is all-or-nothing. Most of the relations which matter in survival are, in fact, relations of degree. If we ignore this, we shall be led into quite ill-grounded attitudes and beliefs.

The claim that I have just made – that most of what matters are relations of degree – I have yet to support. Wiggins' case shows only that these relations need not be one–one. The merit of the case is not that it shows this in particular, but that it makes the first break between what matters and identity. The belief that identity *is* what matters is hard to overcome. This is shown in most discussions of the problem cases which actually occur: cases,

[9] Cf. Sydney Shoemaker, in *Perception and Personal Identity: Proceedings of the 1967 Oberlin Colloquium in Philosophy, loc. cit.*

[10] Cf. Wiggins, p.54.

say, of amnesia or of brain damage. Once Wiggins' case has made one breach in this belief, the rest should be easier to remove.[11]

f→ To turn to a recent debate: most of the relations which matter can be provisionally referred to under the heading "psychological continuity" (which includes causal continuity). My claim is thus that we use the language of personal identity in order to imply such continuity. This is close to the view that psychological continuity provides a criterion of identity.

Williams has attacked this view with the following argument. Identity is a one–one relation. So any criterion of identity must appeal to a relation which is logically one–one. Psychological continuity is not logically one–one. So it cannot provide a criterion.[12]

Some writers have replied that it is enough if the relation appealed to is always in fact one–one.[13]

I suggest a slightly different reply. Psychological continuity is a ground for speaking of identity when it is one–one.

If psychological continuity took a one-many or branching form, we should need, I have argued, to abandon the language of identity. So this possibility would not count against this view.

We can make a stronger claim. This possibility would count in its favor.

The view might be defended as follows. Judgments of personal identity have great importance. What gives them their importance is the fact that they imply psychological continuity. This is why, whenever there is such continuity, we ought, if we can, to imply it by making a judgment of identity.

If psychological continuity took a branching form, no coherent set of judgments of identity could correspond to, and thus be used to imply, the branching form of this relation. But what we ought to do, in such a case, is take the importance which would attach to a judgment of identity and attach this importance directly to each limb of the branching relations. So this case helps to show that judgments of personal identity do derive their importance from the fact that they imply psychological continuity. It helps to show that when we can, usefully, speak of identity, this relation is our ground.

[11] Bernard Williams' "The Self and the Future," *Philosophical Review*, LXXIX (1970), 161–180, is relevant here. He asks the question "Shall I survive?" in a range of problem cases, and he shows how natural it is to believe (1) that this question must have an answer, (2) that the answer must be all-or-nothing, and (3) that there is a "risk" of our reaching the *wrong* answer. Because these beliefs are so natural, we should need in undermining them to discuss their causes. These, I think, can be found in the ways in which we misinterpret what it is to remember (cf. Sec. III below) and to anticipate (cf. Williams' "Imagination and the Self," *Proceedings of the British Academy*, LII [1966], 105–124); and also in the way in which certain features of our egoistic concern – e.g., that it is simple, and applies to all imaginable cases – are "projected" onto its object. (For another relevant discussion, see Terence Penelhum's *Survival and Disembodied Existence* [London, 1970], final chapters.)

[12] "Personal Identity and Individuation," *Proceedings of the Aristotelian Society*, LVII (1956–1957), 229–253; also *Analysis*, 21 (1960–1961), 43–48.

[13] J. M. Shorter, "More about Bodily Continuity and Personal Identity," *Analysis*, 22 (1961–1962), 79–85; and Mrs. J. M. R. Jack (unpublished), who requires that this truth be embedded in a causal theory.

g→ This argument appeals to a principle which Williams put forward.[14] The principle is that an important judgment should be asserted and denied only on importantly different grounds.

Williams applied this principle to a case in which one man is psychologically continuous with the dead Guy Fawkes, and a case in which two men are. His argument was this. If we treat psychological continuity as a sufficient ground for speaking of identity, we shall say that the one men is Guy Fawkes. But we could not say that the two men are, although we should have the same ground. This disobeys the principle. The remedy is to deny that the one man is Guy Fawkes, to insist that sameness of the body is necessary for identity.

Williams' principle can yield a different answer. Suppose we regard psychological continuity as more important than sameness of the body.[15] And suppose that the one man really is psychologically (and causally) continuous with Guy Fawkes. If he is, it would disobey the principle to deny that he is Guy Fawkes, for we have the same important ground as in a normal case of identity. In the case of the two men, we again have the same important ground. So we ought to take the importance from the judgment of identity and attach it directly to this ground. We ought to say, as in Wiggins' case, that each limb of the branching relation is as good as survival. This obeys the principle.

To sum up these remarks: even if psychological continuity is neither logically, nor always in fact, one–one, it can provide a criterion of identity. For this can appeal to the relation of *non-branching* psychological continuity, which is logically one–one.[16]

The criterion might be sketched as follows. ''*X* and *Y* are the same person if they are psychologically continuous and there is no person who is contemporary with either and psychologically continuous with the other.'' We should need to explain what we mean by ''psychologically continuous'' and say how much continuity the criterion requires. We should then, I think, have described a sufficient condition for speaking of identity.[17]

We need to say something more. If we admit that psychological continuity might not be one–one, we need to say what we ought to do if it were not one–one. Otherwise our account would be open to the objections that it is incomplete and arbitrary.[18]

[14] *Analysis*, 21 (1960–1961), 44.

[15] For the reasons given by A. M. Quinton in ''The Soul,'' *Journal of Philosophy*, LIX (1962), 393–409.

[16] Cf. S. Shoemaker, ''Persons and Their Pasts,'' to appear in the *American Philosophical Quarterly*, and ''Wiggins on Identity,'' *Philosophical Review*, LXXIX (1970), 542.

[17] But not a necessary condition, for in the absence of psychological continuity bodily identity might be sufficient.

[18] Cf. Bernard Williams, ''Personal Identity and Individuation,'' *Proceedings of the Aristotelian Society*, LVII (1956–1957), 240–241, and *Analysis*, 21 (1960–1961), 44; and also Wiggins, *op. cit.*, p. 38: ''if coincidence under [the concept] *f* is to be *genuinely* sufficient we must not withhold identity . . . simply because transitivity is threatened.''

I have suggested that if psychological continuity took a branching form, we ought to speak in a new way, regarding what we describe as having the same significance as identity. This answers these objections.[19]

We can now return to our discussion. We have three remaining aims. One is to suggest a sense of "survive" which does not imply identity. Another is to show that most of what matters in survival are relations of degree. A third is to show that none of these relations needs to be described in a way that presupposes identity.

We can take these aims in the reverse order.

III

The most important particular relation is that involved in memory. This is because it is so easy to believe that its description must refer to identity.[20] This belief about memory is an important cause of the view that personal identity has a special nature. But it has been well discussed by Shoemaker[21] and by Wiggins[22] So we can be brief.

It may be a logical truth that we can only remember our own experiences. But we can frame a new concept for which this is not a logical truth. Let us call this "*q*-memory."

To sketch a definition[23] I am *q*-remembering an experience if (1) I have a belief about a past experience which seems in itself like a memory belief, (2) someone did have such an experience, and (3) my belief is dependent upon this experience in the same way (whatever that is) in which a memory of an experience is dependent upon it.

According to (1) *q*-memories seem like memories. So I *q*-remember *having* experiences.

This may seem to make *q*-memory presuppose identity. One might say, "My apparent memory of *having* an experience is an apparent memory of *my* having an experience. So how could I *q*-remember my having other people's experiences?"

This objection rests on a mistake. When I seem to remember an experience, I do indeed seem to remember *having* it.[24] But it cannot be a part of

[19] Williams produced another objection to the "psychological criterion," that it makes it hard to explain the difference between the concepts of identity and exact similarity (*Analysis*, 21 [1960–1961], 48). But if we include the requirement of causal continuity we avoid this objection (and one of those produced by Wiggins in his note 47).

[20] Those philosophers who have held this belief, from Butler onward, are too numerous to cite.

[21] *Op. cit*.

[22] In a paper on Butler's objection to Locke (not yet published).

[23] I here follow Shoemaker's "quasi-memory." Cf. also Penelhum's "retrocognition," in his article on "Personal Identity," in the *Encyclopedia of Philosophy*, ed. by Paul Edwards.

[24] As Shoemaker put it, I seem to remember the experience "from the inside" (*op. cit*.).

what I seem to remember about this experience that I, the person who now seems to remember it, am the person who had this experience.[25] That I am is something that I automatically assume. (My apparent memories sometimes come to me simply as the belief that *I* had a certain experience.) But it is something that I am justified in assuming only because I do not in fact have *q*-memories of other people's experiences.

Suppose that I did start to have such *q*-memories. If I did, I should cease to assume that my apparent memories must be about my own experiences. I should come to assess an apparent memory by asking two questions: (1) Does it tell me about a past experience? (2) If so, whose?

h→ Moreover (and this is a crucial point) my apparent memories would now come to me *as q*-memories. Consider those of my apparent memories which do come to me simply as beliefs about my past: for example, "I did that." If I knew that I could *q*-remember other people's experiences, these beliefs would come to me in a more guarded form: for example, "Someone – probably I – did that." I might have to work out who it was.

I have suggested that the concept of *q*-memory is coherent. Wiggins' case provides an illustration. The resulting people, in his case, both have apparent memories of living the life of the original person. If they agree that they are not this person, they will have to regard these as only *q*-memories. And when they are asked a question like "Have you heard this music before?" they might have to answer "I am sure that I *q*-remember hearing it. But I am not sure whether I remember hearing it. I am not sure whether it was I who heard it, or the original person."

We can next point out that on our definition every memory is also a *q*-memory. Memories are, simply, *q*-memories of one's own experiences. Since this is so, we could afford now to drop the concept of memory and use in its place the wider concept *q*-memory. If we did, we should describe the relation between an experience and what we now call a "memory" of this experience in a way which does not presuppose that they are had by the same person.[26]

[25] This is what so many writers have overlooked. Cf. Thomas Reid: "My memory testifies not only that this was done, but that it was done by me who now remember it" ("Of Identity," in *Essays on the Intellectual Powers of Man*, ed. by A. D. Woozley [London, 1941], p. 203). This mistake is discussed by A. B. Palma in "Memory and Personal Identity," *Australasian Journal of Philosophy*, 42 (1964), 57.

[26] It is not logically necessary that we only *q*-remember our own experiences. But it might be necessary on other grounds. This possibility is intriguingly explored by Shoemaker in his "Persons and Their Pasts" (*op. cit.*). He shows that *q*-memories can provide a knowledge of the world only if the observations which are *q*-remembered trace out fairly continuous spatiotemporal paths. If the observations which are *q*-remembered traced out a network of frequently interlocking paths, they could not, I think, be usefully ascribed to persisting observers, but would have to be referred to in some more complex way. But in fact the observations which are *q*-remembered trace out single and separate paths; so we can ascribe them to ourselves. In other words, it is epistemologically necessary that the observations which are *q*-remembered should satisfy a certain general condition, one particular form of which allows them to be usefully self-ascribed.

This way of describing this relation has certain merits. It vindicates the ''memory criterion'' of personal identity against the charge of circularity.[27] And it might, I think, help with the problem of other minds.

But we must move on. We can next take the relation between an intention and a later action. It may be a logical truth that we can intend to perform only our own actions. But intentions can be redescribed as *q*-intentions. And one person could *q*-intend to perform another person's actions.

Wiggins' case again provides the illustration. We are supposing that neither of the resulting people is the original person. If so, we shall have to agree that the original person can, before the operation, *q*-intend to perform their actions. He might, for example, *q*-intend, as one of them, to continue his present career, and, as the other, to try something new.[28] (I say ''*q*-intend *as* one of them'' because the phrase ''*q*-intend *that* one of them'' would not convey the directness of the relation which is involved. If I intend that someone else should do something, I cannot get him to do it simply by forming this intention. But if I am the original person, and he is one of the resulting people, I can.)

The phrase ''*q*-intend *as* one of them'' reminds us that we need a sense in which one person can survive as two. But we can first point out that the concepts of *q*-memory and *q*-intention give us our model for the others that we need: thus, a man who can *q*-remember could *q*-recognize, and be a *q*-witness of, what he has never seen; and a man who can *q*-intend could have *q*-ambitions, make *q*-promises, and be *q*-responsible for.

To put this claim in general terms: many different relations are included within, or are a consequence of, psychological continuity. We describe these relations in ways which presuppose the continued existence of one person. But we could describe them in new ways which do not.

This suggests a bolder claim. It might be possible to think of experiences in a wholly ''impersonal'' way. I shall not develop this claim here. What I shall try to describe is a way of thinking of our own identity through time which is more flexible, and less misleading, than the way in which we now think.

This way of thinking will allow for a sense in which one person can survive as two. A more important feature is that it treats survival as a matter of degree.

[27] Cf. Wiggins' paper on Butler's objection to Locke.

[28] There are complications here. He could form *divergent* *q*-intentions only if he could distinguish, in advance, between the resulting people (e.g., as ''the left-hander'' and ''the right-hander''). And he could be confident that such divergent *q*-intentions would be carried out only if he had reason to believe that neither of the resulting people would change their (inherited) mind. Suppose he was torn between duty and desire. He could not solve this dilemma by *q*-intending, as one of the resulting people, to do his duty, and, as the other, to do what he desires. For the one he *q*-intended to do his duty would face the same dilemma.

IV

i→ We must first show the need for this second feature. I shall use two imaginary examples. [. . .] [Parfit next briefly describes a case of 'fusion', where two beings with different psychological characteristics (memories, beliefs, character traits, etc.) are amalgamated into a single being.]

We can now turn to a second example. This is provided by certain imaginary beings. These beings are just like ourselves except that they reproduce by a process of natural division.

We can illustrate the histories of these imagined beings with the aid of a diagram. (Figure 1.) The lines on the diagram represent the spatiotemporal paths which would be traced out by the bodies of these beings. We can call each single line (like the double line) a "branch"; and we can call the whole structure a "tree." And let us suppose that each "branch" corresponds to what is thought of as the life of one individual. These individuals are referred to as "A," "B + 1," and so forth.

Now, each single division is an instance of Wiggins' case. So *A*'s relation to both *B* + 1 and *B* + 2 is just as good as survival. But what of A's relation to B + 30?

j→ I said earlier that what matters in survival could be provisionally referred to as "psychological continuity." I must now distinguish this relation from another, which I shall call "psychological connectedness."

Let us say that the relation between a *q*-memory and the experience *q*-remembered is a "direct" relation. Another "direct" relation is that which holds between a *q*-intention and the *q*-intended action. A third is that which holds between different expressions of some lasting *q*-characteristic.

"Psychological connectedness," as I define it, requires the holding of these direct psychological relations. "Connectedness" is not transitive, since these

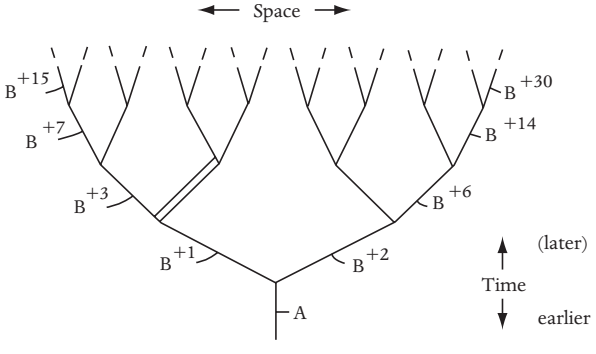

Figure 1

relations are not transitive. Thus, if *X* *q*-remembers most of *Y*'s life, and *Y* *q*-remembers most of *Z*'s life, it does not follow that *X* *q*-remembers most of *Z*'s life. And if *X* carries out the *q*-intentions of *Y*, and *Y* carries out the *q*-intentions of *Z*, it does not follow that *X* carries out the *q*-intentions of *Z*.

''Psychological continuity,'' in contrast, only requires overlapping chains of direct psychological relations. So ''continuity'' *is* transitive.

To return to our diagram. *A* *is* psychologically continuous with *B* + 30. There are between the two continuous chains of overlapping relations. Thus, *A* has *q*-intentional control over *B* + 2, *B* + 2 has *q*-intentional control over *B* + 6, and so on up to *B* + 30. Or *B* + 30 can *q*-remember the life of *B* + 14, *B* + 14 can *q*-remember the life of *B* + 6, and so on back to *A*.[29]

A, however, need *not* be psychologically connected to *B* + 30. Connectedness requires direct relations. And if these beings are like us, *A* cannot stand in such relations to every individual in his indefinitely long ''tree.'' *Q*-memories will weaken with the passage of time, and then fade away. *Q*-ambitions, once fulfilled, will be replaced by others. *Q*-characteristics will gradually change. In general, *A* stands in fewer and fewer direct psychological relations to an individual in his ''tree'' the more remote that individual is. And if the individual is (like *B* + 30) sufficiently remote, there may be between the two *no* direct psychological relations.

k→ Now that we have distinguished the general relations of psychological continuity and psychological connectedness, I suggest that connectedness is a more important element in survival. As a claim about our own survival, this would need more arguments than I have space to give. But it seems clearly true for my imagined beings. *A* is as close psychologically to *B* + 1 as I today am to myself tomorrow. *A* is as distant from *B* + 30 as I am from my great-great-grandson.

Even if connectedness is not more important than continuity, the fact that one of these is a relation of degree is enough to show that what matters in survival can have degrees. And in any case the two relations are quite different. So our imagined beings would need a way of thinking in which this difference is recognized.

V

What I propose is this.

l→ First, *A* can think of any individual, anywhere in his ''tree,'' as ''a descendant self.'' This phrase implies psychological continuity. Similarly, any later individual can think of any earlier individual on the single path[30] which connects him to *A* as ''an ancestral self.''

[29] The chain of continuity must run in one direction of time. *B* + 2 is not, in the sense I intend, psychologically continuous with *B* + 1.

[30] Cf. Wiggins, *op. cit*.

Since psychological continuity is transitive, ''being an ancestral self of'' and ''being a descendant self of'' are also transitive.

To imply psychological connectedness I suggest the phrases ''one of my future selves'' and ''one of my past selves.''

These are the phrases with which we can describe Wiggins' case. For having past and future selves is, what we needed, a way of continuing to exist which does not imply identity through time. The original person does, in this sense, survive Wiggins' operation: the two resulting people are his later selves. And they can each refer to him as ''my past self.'' (They can share a past self without being the same self as each other.)

Since psychological connectedness is not transitive, and is a matter of degree, the relations ''being a past self of'' and ''being a future self of'' should themselves be treated as relations of degree. We allow for this series of descriptions: ''my most recent self,'' ''one of my earlier selves,'' ''one of my distant selves,'' ''hardly one of *my* past selves (I can only *q*-remember a few of his experiences),'' and, finally, ''not in any way one of *my* past selves – just an ancestral self.'' [. . .] [Parfit next briefly describes a second kind of being, which alternately divides into two beings (as in the divided brain case) and then fuses with another beings (as in the case of 'fusion') every six months.]

m→ But let us look, finally, at a third kind of being.

In this world there is neither division nor union. There are a number of everlasting bodies, which gradually change in appearance. And direct psychological relations, as before, hold only over limited periods of time. This can be illustrated with a third diagram (figure 2). In this diagram the two shadings represent the degrees of psychological connectedness to their two central points.

These beings could not use the way of thinking that we have proposed. Since there is no branching of psychological continuity, they would have to

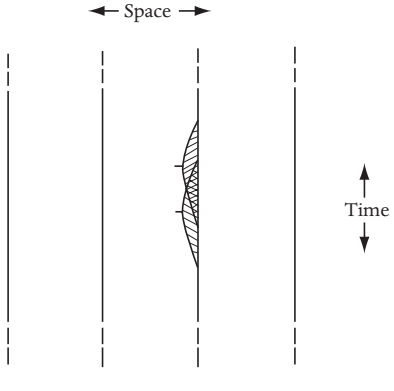

Figure 2

regard themselves as immortal. It might be said that this is what they are. But there is, I suggest, a better description.

Our beings would have one reason for thinking of themselves as immortal. The parts of each ''line'' are all psychologically continuous. But the parts of each ''line'' are not all psychologically connected. Direct psychological relations hold only between those parts which are close to each other in time. This gives our beings a reason for *not* thinking of each ''line'' as corresponding to one single life. For if they did, they would have no way of implying these direct relations. When a speaker says, for example, ''I spent a period doing such and such,'' his hearers would not be entitled to assume that the speaker has any memories of this period, that his character then and now are in any way similar, that he is now carrying out any of the plans or intentions which he then had, and so forth. Because the word ''I'' would carry none of these implications, it would not have for these ''immortal'' beings the usefulness which it has for us.[31]

To gain a better way of thinking, we must revise the way of thinking that we proposed above. The revision is this. The distinction between successive selves can be made by reference, not to the branching of psychological continuity, but to the degrees of psychological connectedness. Since this connectedness is a matter of degree, the drawing of these distinctions can be left to the choice of the speaker and be allowed to vary from context to context.

On this way of thinking, the word ''I'' can be used to imply the greatest degree of psychological connectedness. When the connections are reduced, when there has been any marked change of character or style of life, or any marked loss of memory, our imagined beings would say, ''It was not I who did that, but an earlier self.'' They could then describe in what ways, and to what degree, they are related to this earlier self.

n→ This revised way of thinking would suit not only our ''immortal'' beings. It is also the way in which we ourselves could think about our lives. And it is, I suggest, surprisingly natural.

One of its features, the distinction between successive selves, has already been used by several writers. To give an example, from Proust: ''we are incapable, while we are in love, of acting as fit predecessors of the next persons who, when we are in love no longer, we shall presently have become. . . .''[32]

Although Proust distinguished between successive selves, he still thought of one person as being these different selves. This we would not do on the way of thinking that I propose. If I say, ''It will not be me, but one of my future selves,'' I do not imply that I will be that future self. He is one of

[31] Cf. Austin Duncan Jones, ''Man's Mortality,'' *Analysis*, 28 (1967–1968), 65–70.

[32] *Within a Budding Grove* (London, 1949), I, 226 (my own translation).

my later selves, and I am one of his earlier selves. There is no underlying person who we both are.

To point out another feature of this way of thinking. When I say, ''There is no person who we both are,'' I am only giving my decision. Another person could say, ''It will be you,'' thus deciding differently. There is no question of either of these decisions being a mistake. Whether to say ''I,'' or ''one of my future selves,'' or ''a descendant self'' is entirely a matter of choice. The matter of fact, which must be agreed, is only whether the disjunction applies. (The question ''Are *X* and *Y* the same person?'' thus becomes ''Is *X at least* an ancestral [or descendant] self of *Y*?'')

VI

O→ I have tried to show that what matters in the continued existence of a person are, for the most part, relations of degree. And I have proposed a way of thinking in which this would be recognized.

I shall end by suggesting two consequences and asking one question.

It is sometimes thought to be especially rational to act in our own best interests. But I suggest that the principle of self-interest has no force. There are only two genuine competitors in this particular field. One is the principle of biased rationality: do what will best achieve what you actually want. The other is the principle of impartiality: do what is in the best interests of everyone concerned.

The apparent force of the principle of self-interest derives, I think, from these two other principles.

The principle of self-interest is normally supported by the principle of biased rationality. This is because most people care about their own future interests.

Suppose that this prop is lacking. Suppose that a man does not care what happens to him in, say, the more distant future. To such a man, the principle of self-interest can only be propped up by an appeal to the principle of impartiality. We must say, ''Even if you don't care, you ought to take what happens to you then equally into account.'' But for this, as a special claim, there seem to me no good arguments. It can only be supported as part of the general claim, ''You ought to take what happens to everyone equally into account.''[33]

The special claim tells a man to grant an *equal* weight to all the parts of his future. The argument for this can only be that all the parts of his future are *equally* parts of *his* future. This is true. But it is a truth too superficial to bear the weight of the argument. (To give an analogy: The unity of a nation is, in its nature, a matter of degree. It is therefore only a superficial truth that all of

[33] Cf. Thomas Nagel's *The Possibility of Altruism* (Oxford, 1970), in which the special claim is in effect defended as part of the general claim.

a man's compatriots are *equally* his compatriots. This truth cannot support a good argument for nationalism.)[34]

I have suggested that the principle of self-interest has no strength of its own. If this is so, there is no special problem in the fact that what we ought to do can be against our interests. There is only the general problem that it may not be what we want to do.

The second consequence which I shall mention is implied in the first. Egoism, the fear not of near but of distant death, the regret that so much of one's *only* life should have gone by – these are not, I think, wholly natural or instinctive. They are all strengthened by the beliefs about personal identity which I have been attacking. If we give up these beliefs, they should be weakened.

My final question is this. These emotions are bad, and if we weaken them we gain. But can we achieve this gain without, say, also weakening loyalty to, or love of, other particular selves? As Hume warned, the ''refined reflections which philosophy suggests . . . cannot diminish . . . our vicious passions . . . without diminishing . . . such as are virtuous. They are . . . applicable to all our affections. In vain do we hope to direct their influence only to one side.''[35]

That hope *is* vain. But Hume had another: that more of what is bad depends upon false belief. This is also my hope.

Commentary on Parfit

On first reading Parfit's paper, you might wonder why we should care about the increasingly bizarre thought-experiments he presents. Stories about brains being split in half and housed in two different bodies are pretty far-fetched as it is; but we also have imaginary beings who divide routinely, and even 'immortal' beings. What light can all this possibly shed on the nature of actual, mortal, non-dividing beings like us?

The answer implicit in Parfit's discussion is that it is precisely by examining our intuitions about such cases that we come to a better understanding of issues that really matter to us: whether it is more rational to be concerned with what happens to oneself in the future than to be concerned with what happens to other people, and whether we should find the fact that we will eventually die depressing. Of course,

[34] The unity of a nation we seldom take for more than what it is. This is partly because we often think of nations, not as units, but in a more complex way. If we thought of ourselves in the way that I proposed, we might be less likely to take our own identity for more than what it is. We are, for example, sometimes told, ''It is irrational to act against your own interests. After all, it will be *you* who will regret it.'' To this we could reply, ''No, not me. Not even one of my future selves. Just a descendant self.''

[35] ''The Sceptic,'' in ''Essays Moral, Political and Literary,'' *Hume's Moral and Political Philosophy* (New York, 1959), p. 349.

whether or not Parfit really does succeed in shedding any light on these issues is up for dispute; none the less, the fact that Parfit himself thinks he does succeed should make you think again if your first reaction to the text was, 'how can all this science fiction possibly tell us anything interesting about how things *actually* are?'

Introduction

In the short introductory section, Parfit identifies two very widely-held beliefs which are going to be the targets of his paper. The first is that every question of the form, 'is *X*, at time 1, the same person as *Y* at time 2?' must have an answer. The second is that there are 'important questions' that cannot be answered in the absence of an answer to a question of the first kind.

1. Parfit says at [a]→: 'Our criteria for the identity of [nations and machines] do not cover certain cases'. Say in your own words what he means by this, and illustrate the claim with as realistic and convincing an example as you can. (It doesn't have to involve a nation or a machine; a house, say, would do just as well.)

Section I

Parfit starts by claiming, at [b]→, that if my brain were transplanted into someone else's body, the 'resulting person' would be *me*, on the grounds that they would have 'my character and apparent memories of my life'. He is thus clearly presupposing that our criteria for identity across time for persons are *psychological* criteria.

Parfit next considers the divided-brain case (problem 3 in the introduction to this chapter). His argument runs roughly as follows. Suppose that the question, 'which, out of *B* and *C* (each of whom wakes up after the operation with half of *A*'s brain) is the same person as *A*?', has an answer. Then the answer must be one of these:

1) Neither: *A* ceases to exist, and two new persons (*B* and *C*) come into existence, neither of whom is *A*.
2) Exactly one of *B* and *C*. (Either *B* is the same person as *A*, and *C* is not; or *C* is the same person as *A*, and *B* is not.)
3) The 'sum' of *B* and *C*. After the operation there is really one person (*A*) with two bodies. So *B* and *C* are not in fact two distinct *people* at all.

Parfit spends most of section I arguing that none of the three possible answers is plausible.

2. How exactly does Parfit's discussion at [c]→ of the maths exam fit into this argument?
3. At [d]→, Parfit says that 'we need to prize apart' survival and identity. Explain in your own words what he means and why he thinks this is the moral that ought to be drawn from the divided-brain case.

Section II

The central thesis which Parfit wants us to agree with so far is the claim that person *A* at time 1 can in principle 'survive as' two *different* people – *B* and *C* – at time 2. Clearly, Parfit thinks that 'survival' is an important notion; and, equally clearly, survival has something to do with *psychological continuity*.

The central distinction you need to grasp in order to make sense of this section is the distinction between a relation that is 'logically one–one' and a relation that is 'in fact one–one'. As we have already seen, identity is logically one–one. If person *A* at time 1 is the same person as *B* at time 2, and is also the same person as *C* at time 2, it follows that *B* and *C* are the same person: we have one person at time 2, and not two. We have also seen, from the divided-brain case, that Parfit thinks that *survival* is *not* logically one–one: *A* can 'survive as' two *different* people, *B* and *C*. While *B* and *C* are different people, it would be very odd to say that *A* has ceased to exist, since *B* and *C* each have *A*'s character traits and apparent memories; hence *A* survives as, but is not identical with, each of *B* and *C*.

However, as Parfit says at [e]→, Wiggins' case is 'unlikely to occur. The relations which matter are, in fact, one–one'. The relations which matter, for Parfit, are evidently psychological relations: what matters is *psychological continuity*. Since *in fact* nobody ever gets their brain split in half and housed in two new bodies, we are never actually faced with a situation where one person at an earlier time gets to be psychologically continuous with two different people at a later time; they only ever get to bear the relevant psychological relations to *one* person at a later time. So the relations that matter are *in fact* one–one, but not (as Parfit thinks the divided-brain case demonstrates) *logically* one–one.

One of Parfit's main claims in this section is, in effect, that 'we use the language of personal identity in order to imply [psychological] continuity' ([f]→). When we say that person *A* at time 1 is identical with person *B* at time 2, we imply (*inter alia*) that *B* is psychologically continuous with *B*. Because *in fact* we are never confronted with cases of 'branching' psychological continuity, as with the split-brain case, the language of identity serves us pretty well; we are not forced into absurd consequences, as we would be if split-brain operations were freely available on the National Health Service. But we should not be misled by this into thinking that it is really *identity*, as opposed to psychological continuity, that is what is important.

The other central claim of this section is that the split-brain case should not drive us towards the view that actually it is sameness of body, rather than psychological continuity, that is what is really important.

5. At [g]→, Parfit describes a principle put forward by Williams. (i) Describe in your own words exactly how the principle applies in the divided-brain case (see the paragraph immediately following the statement of the principle). (ii) Describe the two different conclusions that Williams and Parfit draw from its application. (iii) Which conclusion seems most plausible to you? Why?

Section III

In this section, Parfit aims to establish that 'psychological relations' can be understood in a way that does not appeal to facts about personal identity. The objection he is aiming to meet is the 'circularity objection': problem 2 in the introduction to this chapter. Let's first get clear on what he means by 'psychological relations'. Imagine that, last weekend, you formed the intention to go to Paris this weekend; and imagine that you did, in fact, go to Paris this weekend. You thereby *realized* your intention to go to Paris: there is a relation between your intending to go and your actually going. Or imagine that you are now remembering going shopping last Saturday. Then there is a relation between your current state of mind – what we might call your 'memory-experience' – and your previous experience of going shopping.

We have already seen what the circularity objection amounts to in the case of memory: it seems to be a conceptual truth that you can only remember your own experiences, and so we cannot appeal to memory to fix facts about personal identity. The same point applies to other psychological relations – intentions, for example. If *you* intend to go to Paris, then it is a conceptual truth that only you can realize that intention: nobody else's going to Paris could possibly count as realizing or carrying out *your* intention to go.

Parfit attempts to meet the circularity objection by defining psychological relations whose obtaining does not conceptually require that the persons standing, as it were, on each end of the relation are the same person, and by claiming that it is the obtaining of *these* relations that one can appeal to in one's account of personal identity – or (more importantly for Parfit) in one's account of survival.

Parfit's response to the circularity objection is the focus of Schechtman's paper, so we shall return to it in the next commentary; for now, we'll just give an illustration of the kind of thing Parfit has in mind.

Imagine that, far in the future, a company, Rekall, offers 'memory implants' to its customers. (We have borrowed this from the Philip K. Dick short story, 'We Can Remember It for You Wholesale', later made into the film *Total Recall*.) It can be pleasant, on a cold, rainy night, to reminisce about that wonderful beach holiday you had in the Bahamas, when you snorkelled in the warm sea, sunbathed, drank cocktails under tropical palm trees while the sun set, and so on. But what if you could somehow enjoy just the same 'memories' without having gone to the expense of actually taking the holiday? This is the service Rekall offers: 'memories' of holidays and other happy events that you never actually experienced are implanted into your brain.

Now imagine an extra twist. Imagine that it is prohibitively expensive – or perhaps just impossible – to actually *invent* all these apparent memories from scratch. It's much cheaper for the company to take someone else's *actual* memories – someone who actually had a very nice holiday in the Bahamas – and transfer them from their mind into the customers' minds. So when you 'remember' the holiday, what is going

on in your mind – your memory-experience – is exactly like what is going on in the mind of the person who actually had the holiday when she remembers the holiday. But of course you do not genuinely *remember* the holiday, because the experiences you are 'remembering' were not yours but somebody else's.

Imagine that you become hooked on memory-implants: you keep going back for more and more of them, so that eventually you start losing track of which of your apparent memories are really memories and which aren't. By 'apparent memories' we here mean just what Parfit (following Sydney Shoemaker) means by a '*q*-memory' ('*q*' for 'quasi'). An apparent memory *might* be a real memory – you might be recalling an experience that *you* originally had. Or it might not be: it might have been implanted, and somebody else originally had the experience. If you did start to lose track, Parfit thinks you would be in the position described at h→. You might have to work out – by consulting an old diary, say – whether or not a particular *q*-memory is a *q*-memory of your own experience, or of somebody else's.

To sum up: Parfit thinks that psychological relations – *q*-remembering, *q*-intending, and so on – can be used to characterize, non-circularly, a notion of psychological continuity. And we can then use that notion of psychological continuity to characterize survival (or, if we wanted to, personal identity – though Parfit takes himself to have given good reasons already not to want to do that.)

6. Do you think that the story about 'memory'-implants (or rather, *q*-memory implants) is coherent? If not, why not? If you do think it's coherent, do you think *any* memory whatsoever could in principle be implanted into someone else, or could only certain kinds of memory be implanted?

Section IV

In this section, we finally get to understand what Parfit's talk about 'relations of degree' is all about. In this section, his aim – as stated at i→ – is to show that we *need* to be able to think of survival as a matter of degree, and he demonstrates this by using an imaginary example involving 'branching' beings.

7. Explain how the 'branching beings' example is supposed to show that survival can be a matter of degree.

At j→, Parfit draws an important distinction between 'psychological continuity' and 'psychological connectedness'.

8. (i) Explain, in the light of problem 1 described in the introduction to this chapter, why Parfit needs to draw this distinction. (ii) Parfit says at k→ that, for the branching beings, 'connectedness is a more important element in survival' than is continuity. Do you agree? Why or why not?

Section V

Parfit describes a different kind of imaginary being in this section: the immortal beings. (In the original version of the paper, there are also two other examples: 'fusing' beings in section IV, fusing-and-dividing beings in section V.) By this stage, you might be wondering what the point of all this is. Parfit is no longer arguing, or at least not obviously, for or against a particular identifiable philosophical position. So what *is* he doing?

Well, what he is interested in, in these admittedly bizarre cases, is what kinds of ways of thinking and talking the imagined beings might need to adopt in order to express things that are important to them. For example, the dividing beings cannot use the language of identity (recall the moral of section I); Parfit suggests at [l]→ that they might instead use expressions like 'an ancestral self of mine' to denote continuity, and 'a past self of mine' to denote connectedness. Parfit's overall purpose (we think) is to wean us off our addiction to thinking that there is something inherently special about *identity*. Unlike dividing beings, personal identity talk is *coherent* for us, over long periods of time – indeed over a whole human lifetime – because we do not divide. But it doesn't follow that thinking in terms of personal identity is the *best* way for us to think about our own case.

The case of the 'immortal' beings (at [m]→) illustrates this point. Such a being (call him '*S*') *can* think of himself as immortal (just as *we* can think of him as immortal). He can coherently use the word 'I' to refer to someone (call him '*T*') who existed hundreds of years ago but who is psychologically continuous with *S*. But Parfit thinks this would not be a very 'useful' way in which to think. (This is why 'immortal' is in scare quotes: Parfit thinks the beings would *not* think of themselves as immortal.) *S* cannot *q*-remember any of *T*'s experiences, for example. *S*'s attitude to *T* would presumably be one of detached interest – much like our attitude to the past actions of our own great-grandparents, say. We are especially interested in what they (as opposed to other people's great-grandparents) got up to because they are *our* ancestors, but we do not attempt to carry out their intentions, we do not feel remorse for their actions, and so on.

At [n]→, Parfit suggests that we could think of our own lives in much the same way as (he claims) the 'immortal' beings would think about *their* lives – indeed, it would be 'surprisingly natural'.

9. Imagine that human beings only had a memory-span of five years. (i) What kinds of differences (for example, practical, emotional, moral differences), if any, would that make to us? (ii) Do you think we would then regard connectedness (which comes in degrees) as more important than continuity (which is all-or-nothing)? (iii) Do you think we would regard the language of personal identity as just as important as most of us actually regard it? (iv) Now thinking about how we *actually* are, do you think Parfit is right to say that it is 'surprisingly natural' for us to think about our lives in the way that he claims the 'immortal' beings would think about their lives?

Section VI

Parfit finally returns to a claim made at the beginning of the paper: that belief in the special nature of personal identity has bad effects. We would be better off if we thought of 'what matters in the continued existence of a person' as 'relations of degree' [O]→.

10. (i) Describe in your own words the two consequences which Parfit thinks this new way of thinking would have. (ii) Do you think there would be any consequences of this way of thinking that might be bad for us? (iii) Parfit ends the paper with a question. What is he worried about? What do you think is the answer to the question?

Marya Schechtman, 'Personhood and Personal Identity'*

The question 'Who am I?' might be asked either by an amnesia victim or by a confused adolescent, and requires a different answer in each of these contexts. In the former case, the questioner is asking which history her life is a continuation of, and, in the latter, the questioner presumably knows her history but is asking which of the beliefs, values, and desires that she seems to have are truly her own, expressive of who she is. These can be called, respectively, the *question of reidentification* and the *question of self-knowledge*. Contemporary philosophical discussion of what has been called the problem of personal identity is generally considered to be concerned with the question of reidentification. Those involved in this discussion say virtually nothing about the self-knowledge question. Instead, they attempt to spell out the necessary and sufficient conditions for saying that a person at time t_1 is the same person as a person at time t_2 – to give criteria of personal identity over time which would enable us to answer the amnesiac's question.

The primary contenders for a criterion of personal identity have been the bodily criterion and the psychological criterion, which are based, respectively, on the intuitions that it is sameness of body and sameness of personality which are responsible for sameness of person. Of these two, the psychological criterion has been by far the more widely accepted, and

*At all stages of this paper, I was greatly helped by discussion with friends and colleagues. I would like especially to thank Stanley Cavell, Juliet Floyd, Hannah Ginsborg, Dorothy Grover, Anil Gupta, Nick Pappas, Tim Scanlon, Miriam Solomon, Paul Teller, Jennifer Whiting, and the editors of the Journal.

current philosophical discussion of personal identity has focused almost exclusively on attempts to refine and defend this criterion.

In what follows, I shall argue that psychological continuity does not and cannot provide the sort of criterion of personal identity which identity theorists wish to provide. My claim will be that such criteria are inherently circular – that they cannot answer the circularity objection that is standardly raised against them in the literature, and hence cannot provide an analysis of our concept of the persistence of a person.

I shall start, in section I, by giving an exposition of the central features of these psychological-continuity theories. To do this, I shall look in detail at Derek Parfit's[1] version of this theory, and shall use Parfit as a representative of psychological-continuity theorists throughout the paper. I choose Parfit for three reasons. First, his view is, for my purposes, perfectly representative. There is no argument I make against Parfit which could not be applied to any other standard psychological-continuity theory without significant alteration. Second, Parfit's view is one of the strongest versions of this theory, and, finally, it is the version of the psychological-continuity theory which has been most discussed in the recent literature.

After laying out Parfit's view in section I, I proceed, in section II, to describe the circularity objection that has standardly been raised against such views, and the standard response as offered by Parfit. In section II, I shall argue that this response relies implicitly on a highly implausible view of human experience, and that when this view is made explicit it becomes clear that not only does the standard response fail to overcome the circularity objection, but that this objection cannot be overcome, that psychological accounts of identity, to be accurate, must be circular. Finally, in section IV, I shall suggest that the dead end that psychological-continuity theorists have encountered is a result of their conflating the two questions of personal identity outlined above – the reidentification and self-knowledge questions – and suggest what I believe to be a more fruitful direction for philosophical work on personal identity to take.

I

Parfit begins the statement of his version of the psychological-continuity theory with some preliminary definitions. The first is of *direct psychological connections*, which are connections of the following sort: that between a memory and the experience of which it is a memory, that between an intention and the later act in which it is carried out, and the persistence of a belief, desire, or other psychological feature (205–6). Parfit goes on to define a second relation:

[1] *Reasons and Persons* (New York: Oxford, 1984); page references in the text are to this book.

a → *''Psychological connectedness* is the holding of particular direct psychological connections.'' He then says that ''we can claim that there is enough connectedness [for personal identity] if the number of connections over any day, is *at least half* the number of direct connections that hold, over every day, in the lives of nearly every actual person. When there are enough direct connections, there is what I call *strong* connectedness'' (206). He thus takes as the criterion of personal identity over time,

> *The Psychological Criterion:* (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) there does not exist a different person who is also psychologically continuous with Y. (5) Personal identity over time just consists in the holding of facts like (2) to (4). (207)

It is worth considering just what information this criterion is supposed to provide.

Identity theorists take it for granted that we can, in ordinary circumstances, make accurate judgments of personal identity. What they wish to determine is what it is which underlies these judgments. Parfit says that, by 'the criterion of personal identity over time', he means *''what this identity necessarily involves, or consists in''* (203). These theorists are trying to give us an analysis of our concept of personal identity, and so the criterion that they provide will not only have to capture our intuitive notion of personal
b → identity, but it will have to do so in terms more basic. Parfit acknowledges this when he argues that the physical and psychological criteria as he has defined them,

> . . . are both *Reductionist*. They are both Reductionist because they claim (1) that the fact of a person's identity over time just consists in the holding of certain more particular facts,
>
> and
>
> (2) that these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an *impersonal* way. (210)

Parfit is attempting to give an identity criterion that does not employ facts about persons or their identities in its specification, because if such facts were included then this criterion would not provide an analysis of the appropriate sort. It is in the context of this understanding of the project of giving an identity criterion that the circularity objection that I shall discuss in the next section has its force.

II

The circularity objection as it is standardly raised is an adaptation of an objection that was originally raised against John Locke by Bishop Butler. Butler's objection is roughly as follows: memory connectedness seems like a plausible criterion of personal identity, because, properly speaking, we can remember only our own experiences. We can *seem* to remember experiences which are not ours, however, and such seeming memories are no basis for claims of personal identity. A madman may think he remembers leading the troops at Waterloo, but this does not make him Napoleon. Therefore, if memory connectedness is to be a criterion of personal identity, we must have some way of distinguishing between real and merely apparent memory. But that distinction, the argument continues, is no more than this: real memories are apparent memories in which the person remembering is the person who actually had the experience. The obvious problem with the memory criterion of personal identity, then, is that one must already have a criterion of personal identity in order to define memory. Butler argues that, since the fact of identity is prior to the distinction between real and apparent memory, personal identity cannot be defined in terms of memory connectedness.

Psychological-continuity theorists consider a version of this objection, which addresses more general psychological accounts of identity. Parfit indicates that he takes Butler's objection to apply to his psychological-continuity theory when he says:

> On one interpretation, the objection would be this: 'It is part of our concept of memory that we can remember only *our own* experiences. The continuity of memory therefore presupposes personal identity. The same is therefore true of your Relation R [non-branching psychological continuity]. You claim that personal identity just consists in the holding of Relation R. This must be false if Relation R itself presupposes personal identity'. (220)

To defend his psychological criterion, then, Parfit needs to show that none of the connections that constitute psychological continuity – memory connections, the connections between an intention and the action that carries it out, or between the temporal parts of a persistent psychological feature – presupposes facts about personal identity.

Before looking at Parfit's response to this objection, it is worth considering why these other connections might be taken to presuppose personal identity. The objection holds that delusional memories are distinguished from real memories by the nonidentity of the rememberer with the person who had the remembered experience, and thus that when we talk about a genuine memory connection we have already presupposed the persistence of

a single individual. The case is similar with the connection between intentions and the actions in which they are realized.

It seems to be part of our conception of intention that I can only properly be said to intend actions that I can reasonably believe it is in my power to undertake. If I am sane and rational, I cannot, in ordinary circumstances, intend to leap tall buildings in a single bound, though I may fantasize about doing so. Similarly, I cannot intend the actions of other people, because they are not under my control. An intention, then, is necessarily an intention that *I* do something, and so it seems that an action will count as the action that carries out an intention only if the same person who forms the intention takes the action. In this way, the connection between an intention and the action in which it is realized seems to presuppose the persistence of a single person.

We do sometimes speak as if we could frame intentions for others. I may, having asked someone to bring me a book, and finding that she has brought the wrong book, say something like, 'That's not the one I intended for you to bring'. Here, I think, this phrase is an abbreviation for 'That's not the one I intended to direct you to bring', and so the intention in question is still an intention that *I* do something. Even if we do not take such phrases to be abbreviations, however, and do admit that we can form intentions for others, it is clear that the only connections between intentions and actions which are going to be relevant to identity over time are those in which I intend that *I* do something, and *I* carry the intention out, and so the connections that concern us between intentions and the actions that carry them out seem to presuppose the persistence of a single person.

C→ The connections between the different temporal parts of a persistent belief, desire, or other psychological feature seem to presuppose identity in a slightly different way. The problem here seems to be that our saying that we have the same belief or desire at two different times requires not only that there are occurrent beliefs or desires at these two times with the same content, but also that these two occurrent beliefs or desires be had by the same person. Countless people can have beliefs and desires that could be described as having roughly the same content, but this connection does not seem relevant to questions of personal identity. The fact that I believe we should avoid an all-out nuclear war at all costs makes me, in this respect, more *like* all the other people who hold that belief, but it seems irrelevant to the question of identifying me with them. Certainly the fact that someone now holds this belief and someone in the future holds this belief does not, in the absence of any further information, give me any information as to whether or not these two people should be identified. My continuing belief that we should avoid nuclear war is not, therefore, relevant to my persistence through time only because it guarantees that my later self will hold beliefs *like* my earlier self, but because the claim that the *same* belief has persisted presupposes that I have continued to exist.

Butler's objection, then, can be expanded to an objection against psychological-continuity theories. In the case of each psychological connection, qualitative identity of the contents of consciousness is not enough to capture what is relevant to personal identity. It is not enough that the madman's memory is qualitatively like Napoleon's experience, that an action taken by someone else can be described as qualitatively like the action I intended, or that my fellow peace lover's belief that we should avoid nuclear war is qualitatively like mine. Something more must be added to this qualitative similarity to make genuine psychological connections. This something more seems to be persistence of a single individual through the two temporal parts of the connection, and this makes relation *R* look circular.

Parfit's response to this objection derives from a commonly used strategy in the contemporary literature. To give a noncircular identity criterion he must give an account of the distinction between genuine and merely apparent psychological connections which does not appeal to facts about personal identity. Like other psychological-continuity theorists, he attempts to make out this distinction in terms of the cause of the later state. He starts with memory, and says of Butler's objection:

> To answer this objection, we can define a wider concept, *quasi-memory*. I have an accurate quasi-memory of a past experience if
>
> (1) I seem to remember having an experience,
> (2) *someone* did have this experience,
>
> and
>
> (3) my apparent memory is causally dependent, in the right kind of way, on that past experience. (220)

This notion, he says, will answer the objection by giving us a way to distinguish between the kind of apparent memory that is relevant to personal identity and the kind that is not, without appealing to facts about the persistence of persons. Parfit does not explicitly define quasi desires, quasi beliefs, or quasi intentions, but he does recognize their importance within his view, and suggests that they can be specified in the same way that quasi memory is (226).

In the next section, I shall argue that, though quasi states seem initially like a plausible solution to the circularity objection, on deeper reflection it is by no means clear exactly how we are actually to imagine such states, and that on any understanding these states cannot do the work they are supposed to without implicitly presupposing an unacceptably superficial view of the nature of experience. Before turning to my argument, though, it is necessary to consider in a little more depth just how quasi states are supposed to help defend psychological-continuity theories, and to clarify some of the terminology that will be used in what follows. Here I shall

discuss mostly memory connections, but anything I say can be extended to the other connections that make up psychological continuity as well.

d→ The circularity objection is based on the claim that it is part of the definition of memory that we can only properly be said to remember our own experiences. In acknowledging the need to respond to this objection, identity theorists implicitly accept this definition of memory, and, in offering quasi memory as a response to the objection, they implicitly claim that it is not this fact about memory which makes memory connectedness a plausible criterion of personal identity. If, as they claim, quasi-memory connectedness will do just as well as genuine memory in specifying an identity criterion, then it must be the case that whatever it is about memory which makes it a central part of what constitutes personal identity is also present in quasi memory. It follows, then, that it cannot be the fact that memory presupposes the identity of the rememberer with the person who had the experience that underlies the crucial role that memory is taken to play in the constitution of personal identity.

Part of what is behind this conviction, I think, is a conflation of what I take to be two separate issues: the nondelusionality of a memory, and its relevance to the constitution of personal identity. The circularity objection itself relies on a claim that memory, by definition, presupposes personal identity. The examples that are given to support this claim, however, rely on the need to use personal identity to distinguish between delusional and nondelusional memories (e.g., between the madman's apparent memory of Waterloo and Napoleon's), and this distinction is a slightly different one.

In defending the use of quasi memory, psychological-continuity theorists present arguments to show that such connections can provide us with a way of capturing the nondelusionality of a genuine memory without the presupposition of personal identity. I shall argue that this does not yet show us that quasi memories capture what is relevant to personal identity in genuine memory, and that this cannot be captured without the presupposition of personal identity. Before making these arguments, though, it is important to have a clear view of how it is that quasi memories are thought to be nondelusional.

The idea, I believe, is this. What makes the madman's apparent memory of Waterloo delusional is not simply the fact that an apparent memory with this qualitative content occurs in his psyche, but rather the fact that he takes this apparent memory to be of an experience that he had. To just have an apparent memory with this content and remain agnostic about whose experience is being remembered would not be to have a delusion on this

e→ view, even if the quasi rememberer has had no experience that corresponds in content to the quasi memory. The distinction that is being made, then, can be described as follows: to have a *memory*, is to have an apparent memory of an experience that one actually had, and to take it (correctly) to be one's own experience. To have a *delusion* is to have an apparent memory of an

experience that one did not in fact have, and to take it (incorrectly) to be one's own. To have a *quasi memory* is to have an apparent memory (properly caused) and to hold no view about whose memory it is.[2] Insofar, then, as this last case describes a nondelusional apparent memory, quasi memory seems to give us a way of specifying nondelusionality without reference to sameness of person.

That this is the kind of categorization that identity theorists have in mind can be seen in the example of quasi memory that Parfit gives. After giving his definition of quasi memory he says: ''We do not quasi-remember other people's past experiences. But we might begin to do so. . . . Suppose that . . . neuro-surgeons develop ways to create in one brain a copy of a memory-trace in another brain. This might enable us to quasi-remember other people's past experiences'' (220). He then goes on to offer an example, the story of Jane and Paul, in order to show that it is possible for one person to have another person's memory nondelusionally. We are to imagine that Jane has copies of some of Paul's memory traces implanted in her brain, and shortly thereafter has a new vivid set of apparent memories of experiences that she knows she never had. In particular, she has an extremely vivid memory of a storm in Venice and of a bolt of lightning hitting the bell tower of San Giorgio, though she has never been to Venice. On asking Paul, she discovers that he did have exactly such an experience. Parfit concludes: ''Given all of this, Jane should not dismiss her apparent memory as a delusion. She ought to conclude that she has an accurate quasi-memory of how this flash of lightning looked to Paul'' (221). Jane, then, has an accurate recreation of the qualitative content of Paul's memory of seeing the flash of lightning, and it is (by Parfit's definition), appropriately caused. This qualitative similarity is to be taken quite seriously; we are really to think of Jane as having Paul's memory exactly recreated, and, in regard to this one incident, experiencing exactly the inner state Paul would in a genuine memory.

Having set up the example, Parfit immediately considers the objection that for Jane to be given Paul's memory in this way would amount to her having a delusion, and he dismisses it. He says: ''It may be claimed: 'Since Jane seems to remember *seeing* the lightning, she seems to remember *herself* seeing the lightning. Her apparent memory may tell her accurately what Paul's experience was like, but it tells her, falsely, that it was *she* who had this experience' '' (221). If this objection stands, then quasi memory will be no help against the circularity objection. But Parfit has a response. He says:

f→ Because we do not have quasi-memories of other people's past experiences, our apparent memories do not merely come to us in the first-person mode. They come with a belief that, unless they are delusions, they are about our

[2] I am extremely grateful to Paul Teller both for suggesting the necessity of laying out this distinction, and for his helpful comments in formulating it.

> own experiences. But in the case of experience-memories, this is a separable belief. If like Jane we had quasi-memories of other people's past experiences, these apparent memories would cease to be automatically combined with this belief. (222)

This, then, supports my claim that quasi memory is thought to avoid the problems of the circularity objection because it can characterize a nondelusional memory without reference to personal identity.

In the remainder of this paper, I shall argue that quasi memory fails to do the work it is supposed to, avoiding the circularity objection because relations between the specification of the qualitative content of a memory, presuppositions about the identity of the rememberer, and the issue of nondelusionality are far more complex than identity theorists recognize. I g→ shall argue that simply deleting the ''nametag'' from a memory is not sufficient to make it nondelusional, and that in order to make an apparent memory truly nondelusional one will either have to presuppose the identity of the rememberer with the person who had the experience, or else remove so much of the content of the memory that it is no longer plausible to say that what is relevant to personal identity in genuine memory is preserved in quasi memory. So I shall argue that not only do quasi states not fulfil the role that they are supposed to, but furthermore that they cannot – that there is no way to capture what is relevant to personal identity in memories without presupposing identity.

III

I shall start my argument with a detailed look at an example offered by Edward Casey in his book, *Remembering: A Phenomenological Study,*[3] where he describes one of his own memories. If we ask what would happen if this memory experience were given as a quasi memory, the limitations of quasi states will become clearer. Casey recalls a memory of a recent family outing:

> I recall going to the movie *Small Change* a few weeks ago – exactly when, I am not certain. After dinner nearby at Clark's, my two young children, my wife, and I had walked briskly over to the Lincoln Theater, stopping briefly at a paperback bookstore on the way. Anticipating a large crowd, we arrived early and were among the first to purchase tickets. There ensued a wait that seemed much longer than the ten or fifteen minutes it actually was. The children were especially restive and had difficulty staying in the line that had formed – Erin attempting some gymnastic tricks on the guardrail by the entrance, Eric looking at the posted list of coming attractions. Finally the doors were flung open, and we entered at the head of what was, by then, a considerable line. Once inside, we

[3] Bloomington: Indiana UP, 1987.

> sought seats approximately in the middle of the theater, settled there, and interchanged positions a couple of times to adjust to the height of those sitting in front of us. The lights dimmed, and *Small Change* began directly. (Or was there not a short feature first? – I cannot say for sure.) The film was in French, with English subtitles. I have only a vague recollection of the spoken words; in fact, I cannot remember any single word or phrase, though I certainly remember the characters *as speaking*. The same indefiniteness applies to the subtitles, at which I furtively glanced when unable to follow the French. Of the music in the film I have no memory at all – indeed, not just of *what* it was but *whether* there was any music at all. In contrast with this, I retain a very vivid visual image of the opening scene, in which a stream of school children are viewed rushing home, seemingly in a downhill direction all the way. Two other scenes also stand out in my present recollection: an infant's fall from the window of a high-rise apartment (the twenty-ninth floor?) and the male teacher (whose name, along with all others in the film, I have forgotten) lecturing passionately to his class about child-abuse. Interspersed between these scenes is a medley of less vividly recalled episodes, ranging from fairly distinct (the actions of a child-abusing mother) to quite indistinct (e.g., children's recitations in the classroom). While I am recollecting this uneven and incomplete sequence of filmic incidents, I find myself at the same time remembering my own children's ongoing reactions to the film. I do not remember their behavior in detail but only as a kind of generalized response consisting of laughing, whispered questions, outright comments, and the like. These reactions are as intrinsic to the memory as is the unfolding of the film itself; so too is the mixture of pleasure and exasperation which I felt in being located, as it were, *between* children and film. Suddenly my memory of *Small Change* comes to an end: the lights go up, and we leave through a side exit near us, overhearing expressions of amusement and satisfaction from those around us as we walk out into the night. (ibid., pp. 25–6)

Parfit's claim is that our relation to the experiences we remember (taking them to be our own) is separable from their content, and that, if this relation is changed, we can be given someone else's memory nondelusionally. There are a number of features of Casey's memory which show this view to be problematic.

One striking difference between this example of an actual memory and the fictitious example that Parfit gives us is the detail and richness of the former as opposed to the latter. Casey's memory of the trip to see *Small Change* includes memories of his family, the restaurant where he has eaten more than once, the theater at which he has seen more than one movie, and so on. Certainly memories that consist of a striking visual image, like the memory of Venice that Parfit describes, are one sort of experience memory, but more of our memories are like the one Casey describes; they are memories of family vacations, or writing a paper, or turning thirty. These memories are much more complex due to the number of personal associations and details about the individual's life which are involved in them. Intuitively, it seems that these memories are more likely to be the sort that

lend plausibility to the idea that memory connectedness is an essential feature of personal identity than memories of the sort Parfit describes, and in any event, since quasi memory is meant to provide a defense of psychological-continuity theories, memories of this richer sort will also have to be possible quasi memories.

This fact, however, commits Parfit and other identity theorists to some quite implausible views about the inner lives of persons. To see this, let us consider what it would be like for Jane to be given Casey's memory of *Small Change* as a quasi memory. The amount of personal detail that this memory includes makes it difficult to imagine Jane receiving it as a quasi memory. Some striking visual image is an experience that could, given the appropriate circumstances (i.e., if Jane had really been in that spot in Venice at that time), be had by virtually any sighted person, regardless of the rest of her psychological makeup. Casey's memory is not like that; it contains a good many elements that make reference to other parts of his life and his personality. It includes, for instance, the familiarity with the town in which he lives, the restaurant where he eats, the theater where he has seen numerous movies, and so on. His knowledge or lack of knowledge of French also plays a role. More important, perhaps, is the fact that his relationship to his children is central in what he remembers; his response to the gymnastics on the guardrail, and his mixture of pleasure and exasperation at their responses to the movie are, he says, essential parts of this memory.

These features of the memory are going to be very alien to Jane, who presumably does not have a wife, and, we may assume to make our point, does not have children. Not only will the physical locations seem unfamiliar, but so much else about the memory will seem anomalous. If Jane speaks fluent French, then the experience of reading the subtitles will be puzzling. And if she happens to have acted in this movie, or to be a musician, then her failure to recognize herself, or to remember whether there was music will be very disturbing. Furthermore, the idea of having a wife and children, as well as the interest/pride/embarrassment/concern relating to the behavior of these particular children, remembered as Erin and Eric, will be out of place in Jane's psyche.

h→ Once we have noticed that this memory contains so many elements personal to Casey and anomalous to Jane, it is no longer easy to imagine what her quasi memory will be like. It seems to me that there are only two possibilities. The first is that Jane will reproduce all of the visual content of the memory without interpreting it as Casey does. That is, upon awakening from the quasi-memory implant surgery, Jane will have images of being in a strange restaurant, then a bookstore, then a theater, with a woman and two children whom she does not recognize, and she will also have images of seeing a movie with these people. The second alternative is that she will reproduce the memory exactly as it occurs in Casey, with all of the same

personal elements and associations. Neither of these alternatives will allow quasi memory to do the work it is supposed to.

Consider the first alternative. It is, first of all, not entirely clear that this alternative is even actually possible. Our memories may not be strictly visual in the way that this alternative suggests. We should not take it for granted that Casey's remembering his family, for instance, is his having visual images of some people whom he is able to recognize. Instead he may remember them nonvisually as his wife and children, and this would be impossible to reproduce neutrally in Jane. Even if there were a way to reproduce this memory in some neutral way in Jane's consciousness, though, this alternative would fail because quasi memory of this sort would not capture what is relevant to personal identity in genuine memory connections. If so much of what this memory is to Casey is missing, if instead of a memory of a family evening out there is only a series of images, then we have little basis to say that Jane's quasi memory is qualitatively the same as Casey's memory, even though it does contain many of the same elements. Since memory is serving as a possible basis for personal identity, it seems that the personal elements, the ones that will be missing in Jane, are likely to be the most crucial.

There is good reason to question whether Jane's experience, so understood, is qualitatively like a memory at all. Phenomenologically, it will be drastically different from Casey's experience, appearing not as a coherent memory of a family outing, but as a blur of unidentifiable sights and sounds, which will make it not only unlike Casey's memory, but unlike any memory. If this quasi memory is different from Casey's experience in such a fundamental way, then we seem to have no good reason to call it a quasi memory of Casey's experience. This in turn makes it seem quite unreasonable to say that this quasi memory captures what is relevant in the connection between a genuine memory and the experience remembered.

The second possible way to imagine this case is to imagine Casey's memory reproduced in Jane exactly, with all of its personal elements. In this quasi memory, then, it will seem to Jane that these people are her wife and children; Clark's, the bookstore, and the Lincoln Theater will seem familiar, and the emotional responses to the movie and to Casey's family will seem like her own. If this is what happens, though, then it is difficult to make out Parfit's claim that the belief that an experience I remember is my experience is separable from the memory itself. If the memory must be such that I think of it as *my* family and *my* hometown, then the mineness of the experience seems to be part of the content of the memory. Parfit might respond that, whereas that is true to a point, the belief is separable at a later stage; whereas the apparent memory itself may carry the semblance of the experience being hers, on reflection Jane would be able to separate this out. This would involve her having a memory qualitatively just like Casey's, but reflecting on it and saying to herself something like, 'These people seem to

be my family, so they must be the family of whoever's memory I've been given, for they certainly aren't mine'.

There are problems with this solution, too, however. The first is that it is again not clear that we are entitled to say that Jane's quasi memory is qualitatively the same as Casey's memory. Although it contains all of the same elements, there are additional elements that will be added by Jane's psyche. If Jane has no children, no spouse, never goes to movies, and has never been in Casey's hometown, then surprise and confusion are likely to accompany this quasi memory in Jane. It will seem extremely foreign, anomalous, and as if it cannot be right but must be some sort of delusion or fantasy. This is not just a memory that Jane cannot quite place; it is one that cannot, in principle, be a part of a coherent life history for her, since it contradicts what she knows about herself. A memory that is familiar and nostalgic to Casey seems foreign and fantastic to Jane. It seems quite plausible, therefore, to say that this memory, with its dimension of strangeness, is a qualitatively different experience for Jane than for Casey. If, therefore, we really wanted to reproduce the qualitative content of Casey's memory in Jane, we would not only have to recreate a great many of Casey's states in Jane, but suppress a great many of Jane's as well, and this begins to look suspiciously like replacing Jane's psychology with Casey's.

Even if we took this route, though, Parfit would still be in trouble because Jane's memory would no longer be nondelusional. We have seen that giving Jane the full content of Casey's memory requires the inclusion of facts about who has the memory. These facts, however, are false in relation to Jane, so their inclusion will make her quasi memory delusional. For Jane to experience something like Casey's memory, she will have to take his family to be her family, his town to be her town, his emotional reactions to be her emotional reactions, and so on. But these are not her family, town, and emotional reactions any more than Napoleon's troops are the madman's, and as long as Jane has to take these components of Casey's memory to be her own, there is no way to make this memory nondelusional.

This discussion has shown that quasi memory cannot do what it was devised to do, because presuppositions about who has a psychological state come in at a level deeper than the level of the connections between states; they are necessary to defining those states as well. The original circularity objection charged that specifying the connections between psychological states required appeal to facts about the identity of the person who had the states. Parfit's original response, therefore, was that he needed to find an impersonal way of differentiating between two qualitatively identical apparent states (say, the madman's memory of Waterloo and Napoleon's), one of which was genuinely connected to an earlier state and one of which was not. It was presumed that the content of the memory could be defined independently of facts about whose memory it was, and that all that was needed to avoid circularity was to show that the relevant connection of this impersonally defined memory to an

impersonally defined earlier experience could also be defined impersonally. This is what Jane's case was supposed to show.

What I have argued, though, is that facts about whose memory a given memory is are an integral part of its qualitative content. Without the associations to facts about the rememberer's life and psychology which a memory ordinarily has, an experience loses so much of its content that it seems arbitrary to call it even the same apparent memory, and so to take it to be in any way relevant to questions about personal identity. If, on the other hand, the entire content of the apparent memory is reproduced, then the inclusion of the facts that allow this will guarantee that the memory is delusional. The fact, then, that presuppositions about who has a memory are inseparable from its content means that one cannot, as Parfit claims, specify nondelusionality impersonally by keeping the content of a memory and simply deleting presuppositions about whose memory it is. . . . [Schechtman next argues briefly that other quasi states – in particular, beliefs and desires – are subject to the same difficulties.]

What this discussion has shown is that, on either of the two possible pictures of what it is to have a quasi state, quasi states fail to do the work they are supposed to because they include either too little or too much of the state they reproduce. If they include too little, they do not capture what is relevant to personal identity, and if they include too much, then, unless sameness of person is assumed, they are delusional.

IV

My objection is not just an objection against quasi states, but also against the project that includes them. The attempt to give a noncircular analysis of identity over time requires that the distinction between genuine and apparent continuity be made without presuppositions about the identity or existence of the person having the states which constitute such continuity. This turns out to be impossible. My argument that quasi states cannot serve this function has not relied on any particular facts about such states as Parfit defines them; it has relied only on the fact that we cannot imagine a state qualitatively like Paul's occurring in Jane without that state being delusional. The mineness of a psychological state cannot be separated from its content, and so, to define a state that can be properly called one of the states constitutive of personal identity, we will have to presuppose the existence of a persistent person who is the subject of that state. This means that the strong noncircularity condition cannot be met, and so an analysis of personal identity of the type identity theorists wish to give is not possible.

It would certainly be unfortunate if this meant that we had to give up on psychological accounts of personal identity altogether, since the intuition that psychology and personality are essential to our identities is extremely

strong. Luckily, the above discussion does not have this consequence. It does, however, require us to rethink what such criteria will be like, and what will be required of them. I would like to conclude, then, by offering a sketch of an alternative way of conceiving of what it is to give a psychological account of identity. This alternative promises to capture many of the intuitions that support Parfit's psychological-continuity theory without running into the problems which his view does. Such a view will not provide us with a reductionist analysis of personal identity of the sort that psychological-continuity theorists try to offer, but it can provide us with a great deal of insight into what is involved in the identities of persons by emphasizing rather than neglecting the complexities of our psyches.

In order to understand this alternative better, it is helpful to return to the issues that I raised at the beginning of this paper. There I suggested that the problems that psychological-continuity theorists encountered could be explained by their conflation of the two questions of personal identity: the question of reidentification, on the one hand, and of self-knowledge, on the other. I cannot fully argue for this claim here, but I can say something about how such an argument would proceed, and this will be helpful in developing an alternative psychological account of personal identity.

j→ It is a philosophical commonplace that we have a dual perspective on persons. On the one hand, we view persons as one of the types of objects in the world, but, on the other, we view them as subjects and agents, creatures with a way of experiencing the world and with affect and volition. Because we have these two views, we tend to use the word 'person' in two different ways. Viewing persons as objects of our knowledge, we tend to use the term as if it were more or less coextensive with 'live human body'. Viewing persons as subjects and agents, we tend to use 'person' as an honorific term, having to do with autonomy, moral agency, and volition. In this sense, not all human beings will necessarily be persons, and not all actions of human bodies will be actions of persons.

k→ It is my claim that the two questions of personal identity come from these two perspectives also. Insofar as we view persons as objects of our knowledge, it will make sense to view the question of personal identity as a subspecies of more general metaphysical questions about the identity of complex objects, and to employ the same methods that are used in this pursuit to answer questions of personal identity. These methods involve looking at time slices of persons from an atemporal third-person perspective, and attempting to give objective criteria for saying that two time-slices belong to the same person. Ultimately, I believe, these methods can only support a bodily criterion of personal identity. Viewing persons as subjects and moral agents, on the other hand, discourages such an objective approach to identity, and leads us to emphasize the subjective experience of a person and questions of responsibility in addressing questions of identity (the sorts of identity questions raised in ethics, action theory, and literature, for example).

My claim is that contemporary psychological-continuity theorists like Parfit have taken their goal of providing a noncircular identity criterion from the reidentification question, and the intuitions that support the psychological criterion over the bodily criterion from the self-knowledge question. It is this, I claim, which has pushed these theorists to hold such an implausible view of our psychological states. The arguments that support a psychological criterion all rely on hypothetical puzzle cases, in which we are asked to imagine someone changed in certain ways, and then to observe our intuitions about whether or not the original person should be concerned about the future of the resultant person, held responsible for her actions, or take her beliefs to be her own. Issues of agency and self-knowledge are emphasized, and it is on the basis of these that we are being asked to judge questions of identity. The assumption is that when we concentrate on such questions we will judge that responsibility and concern (and hence identity) go with the psychology or consciousness rather than the body.

These thought experiments, then, are taken to support the view that psychology rather than the body is the constitutive factor of identity, but the question is never raised as to what kind of thing it is the identity of which is at issue – the person taken as subject or as object. The intuitions we have in response to the thought experiments, I claim, come from our view of persons as subjects, but the methods that identity theorists use to turn these intuitions into a criterion of identity come from a view of persons as objects. Taking the fact that psychology is what turns out to be important in these cases, psychological-continuity theorists thus make the unwarranted assumption that sameness of psychology can be used to provide a noncircular criterion of identity of the sort which is given for objects. But such a criterion cannot focus on subjectivity; it is, by definition, to be objective, and must be capable of being spelled out without including the first-person perspective of a given individual. The pieces that make up a person's psychology, must, to fulfil this purpose, be viewed to be as discrete and detachable as are the planks of a ship or the grains of sand in a heap. It is because psychological-continuity theorists are trying to force the insights gained from consideration of questions of self-knowledge and responsibility into the mold of questions of the persistence of material objects that they are forced to view psychological states as atomic, isolable, and in principle independent of the subject who experiences them – a view that I have argued to be highly implausible.

The intuitions that support psychological criteria come from the considerations that are at play in the self-knowledge question, and so it seems reasonable to hope that the sort of work done to address this question will be of some help in reconceiving the notion of a psychological account of personal identity. Let us consider this question for a moment. The person interested in self-knowledge will ask herself questions of the following sort: 'Is this really what I believe?' 'Is this really what I desire?' 'Is this really what I

intend?' Which is to ask, in a specific way, 'Is this my belief?' 'Is this my desire?' 'Is this my intention?' The resources involved in answering these questions would be resources that allowed a person to assign particular psychological states to herself, and so would be at the same time resources for delimiting an individual person, and for answering questions of personal identity.

What kind of information, then, can be used in answering these questions? We have seen one relevant set of considerations already in the discussion of quasi states. In each of the cases we imagined, we could identify something highly anomalous about the states given to Jane even independently of the knowledge that they had come from someone else. We saw that Casey's memory or Paul's desire to join the French Foreign Legion, put into Jane's psyche in full form, were not only delusional, but jarring. This is because they contained so many details that were in conflict with the details of Jane's own life. Jane, I claimed, would be unable to appropriate these states coherently as her own, because they would conflict so violently with what she knows to be true of herself. If Paul's desire to join the French Foreign Legion were, therefore, to occur in Jane's psyche, she would have good reason to doubt that it was really her desire, even if she had no prior knowledge of its origins.

Part of what is involved in a psychological state's being mine in the sense which is at issue in the self-knowledge question, then, is its coherence with my total psychology – my ability to view it as a comprehensible part of my life, and to take it to be my own. Clearly the sort of coherence required here cannot be strict logical consistency – few of us are fortunate enough to have psyches without conflict, ambivalence, or discrepancies in our memories – but there is some way in which we are able comfortably to appropriate the various aspects of these ordinary conflicts, and view them as part of a single, comprehensible life. When the conflict is too severe, however, and the states in question are incomprehensible to us within the context of our self conception – as Casey's memory or Paul's desire might be to Jane – then these states have the character of external impositions, and a question is raised as to whether they are a genuine part of the person in whom they occur.

It is precisely when someone runs into an anomaly of this sort that we say that she is having an "identity-crisis," and the resolution of this crisis generally involves the forming of a well-grounded self-conception, and the ability to reject as not one's own beliefs, desires, and character traits that are not comprehensible as part of that self-conception.[4] As an alternative approach to offering a psychological account of personal identity, I propose

[4] I cannot emphasize enough that this is the sketch of the view, and that many distinctions and definitions are required to fill it out. What I want to make clear here is that what I have just said does not mean, for instance, that, if I conceive of myself as a patient person, and find suddenly that I have murderous impulses concerning the person driving five miles an hour in front of me on a one-lane road, I

that we take this talk of "identity crises" literally and not metaphorically – that we take persons (viewed as subjects and agents) to be constituted by their own self-conceptions. This view makes sense both in light of the discussion of quasi memory – which shows that the inability of someone to appropriate an experience coherently is sufficient reason to say that it is not hers – and in light of the intuitions that support the psychological criterion in the first place. These intuitions, I argued, are based on a notion of persons as self-conscious agents who are capable of moral responsibility, and such a conception seems to require as an essential feature of persons that they have a coherent self-conception. It is, after all, precisely when we find that persons do not have a coherent conception of themselves as agents that we feel we cannot legally hold them responsible for crimes they have committed – that there is a relevant sense in which these crimes are not their actions.

Furthermore, there is much philosophical precedent for viewing persons as self-constituting, the fact of their identities created by their self-conceptions. This view is central in the work of the existentialists, but it can also be found closer to home in the work, for instance, of Harry Frankfurt,[5] or Herbert Fingarette.[6] It is also very much present in the work of Locke[7] himself, who tells us that a person is "a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places" (ibid., p. 39), and that "Wherever a man finds what he calls himself there, I think, another may say is the same person" (ibid., p. 50). Since contemporary psychological-continuity theorists take themselves to be addressing the question that Locke raises, it does not seem too far-fetched to believe that a view of persons as created by their self-conceptions, which Locke himself proposed, could provide a means of answering this question.

Of course, this view needs a great deal more development. I have already mentioned the need to spell out in more detail both what a self-conception is and what coherence is. The madman with Napoleonic delusions takes himself to have led the troops at Waterloo, but this does not count toward making that his action. And my refusal to accept my competitive impulses as my own does not have the consequence that I am not a competitive person. Also, it seems clear that I cannot use just *any* material to form my

can simply reject these murderous impulses and thereby make them not mine. I understand exactly where these impulses come from, and am forced, in the face of them, to revise my view of myself as a patient person. The type of incomprehensibility that I have in mind here is stronger and of a different sort. I shall say more about this later in the paper.

[5] See, for instance, his "Identification and Externality," in *The Identities of Persons*, Amelie Oksenberg Rorty, ed. (Los Angeles: California UP, 1975).

[6] See, for instance, his *Self-Deception* (London: Routledge & Kegan Paul, 1969).

[7] "Of Identity and Diversity," from *An Essay Concerning Human Understanding*, reprinted in *Personal Identity*, John Perry, ed. (Los Angeles: California UP, 1975).

self-conception. Certain things I must view as part of my life if I am ever to have such a conception, and other things I cannot: there is no coherent way for me to appropriate Napoleon's actions at Waterloo. More must be said, then, about what it is which determines what can be involved in my self-conception, and about what a self-conception is.

Fortunately, we already have many resources available to help in filling in the gaps which we find in this view. There are, first of all, the philosophical sources that I mentioned above. We can also learn a great deal from work in psychology. Developmental psychology tells us how an infant progresses from a creature who does not distinguish between itself and the external world, to a person who distinguishes clearly between what is himself and what is other. And abnormal psychology provides us with analyses of cases where the attempt to form an identity goes awry. There are also many rich suggestions in literary treatment of identity, metamorphosis, and the double, which can be of tremendous assistance in trying to spell out this view. There is good reason to hope, then, that an alternative psychological account of persons of the sort I propose could be developed.

A view of persons as self-constituting would, when developed, provide an identity criterion of sorts. If we wanted to know, for instance, whether Jane is the person who had some particular experience, then we would need to discover whether this experience is a part of her coherent self-conception, because it is this which is the necessary and sufficient condition for assigning an experience to a particular person. This criterion, it is clear, would not fulfil the strong noncircularity condition to which psychological-continuity theorists hold themselves. It is not a reductionist analysis of persons, because it requires that in each instance we settle questions of identity with reference to the subjectivity and psyche of the person in question. I hope that the sketch I gave above will be sufficient to show, however, that this circularity is nonvicious – that such a criterion can provide us with a great deal of insight into what it is which constitutes the identities of persons, and that it can do so without doing violence to the complexities of our experience.

Commentary on Schechtman

Schechtman's paper is in many ways a model philosophy paper. The structure of the overall argument is crystal-clear. The paper contains, in section III, a single objection which strikes at the heart of all psychological-continuity theories of personal identity, the objection being that appeal to 'quasi states' fails to save such theories from the circularity objection. The paper is not merely destructive, however. Schechtman offers, in section IV, a diagnosis of the failure of psychological-continuity theories, and sketches an alternative view about personal identity – a view that is explicitly aimed at preserving the intuitions which lie behind the appeal of such theories without succumbing to the objection she raises.

Notice, however, that getting on for half the paper is devoted to scene-setting. In the short introductory section, Schechtman provides some context for the paper, tells us clearly why she has focused on Parfit in particular, and sketches a plan of the paper. In section I, she briefly lays out what she takes to be Partfit's view, as expressed in his later book, *Reasons and Persons* (Parfit 1984). And in section II, she provides a very clear and thorough explanation of what the classic circularity objection to psychological continuity theories is, and how Parfit's appeal to 'quasi states' is supposed to solve it. This expository material is useful in its own right, but it also makes Schechtman's objection much easier to grasp.

Section I

The first thing to notice, at a→, is that the view of Parfit's – from *Reasons and Persons* – described here is not the same as the view expounded in his 'Personal Identity'. In the latter work, as we have seen, Parfit is using the notion of psychological connectedness in his analysis of *survival*, and arguing that we can and should abandon the language of personal identity in favour of the language of survival. The view described in *Reasons and Persons* that Schechtman is going to criticize, by contrast, is one according to which personal *identity* can be defined by appealing to the *proportion* of 'direct connections' that hold between a person on one day and a person on the next day.

Fortunately the details need not detain us here, because it doesn't actually matter for the purposes of Schechtman's argument whether one thinks of psychological connectedness as grounding personal identity or merely survival; nor do the precise details of *how* one puts the notion of psychological connectedness to work in one's account of personal identity (or survival). For Schechtman's argument is going to be an attack on the very possibility of defining psychological connectedness, and hence continuity, in a way that does not presuppose facts about personal identity.

Next, Schechtman notes that the criteria for personal identity that Parfit proposes are 'reductionist'. Reductionism is an important concept in metaphysics. Roughly, a reductionist view about a certain kind of entity is a view according to which that kind of entity is really nothing over and above some *other* kind of entity. For example, a reductionist about the mind may say that the mind is really nothing over and above the brain: the mind isn't an extra addition to our ontology. A reductionist about goodness and badness may say that the goodness or badness of an act is really just a matter of how much pleasure or pain it causes: if we already believe that certain acts cause certain levels of pain or pleasure, we do not have to believe in anything extra in order to think that they are bad or good. And so on.

1. (i) Look at Parfit's reasons for saying that physical and psychological criteria for personal identity are reductionist, at b→. Think of some other philosophical theories you have come across. Which of them do you think are reductionist? Which are not? In each case, explain your answer. (ii) Why might one think that it is a good thing to offer a reductionist theory of something? Do you agree? If so, why? If not, why not?

Section II

Schechtman starts by describing the circularity objection. She claims that the objection works in different ways for different kinds of psychological state (memories, intentions, beliefs, and so on). In the case of memory, the idea is that in order for a mental state *m* to be a genuine memory of some past experience *e*, the person in state *m* must be the very same person as the one who originally had the experience *e*. Otherwise state *m* is not a memory at all; it is just a delusion. In the case of intention, for me to have a genuine intention to perform some act *A*, it must be an intention that *I* perform act *A*. In general, for person *Y* to carry out the intention of person *X*, it must be the case that person *X* = person *Y*. In the case of belief, in order for person *Y*'s belief that nuclear war should be avoided at all costs to be the *very same belief* as person *X*'s belief that nuclear war should be avoided at all costs, again, it must be the case that *X* = *Y*.

2. Describe in your own words the distinction drawn at [c]→ between (i) *X* and *Y* having a belief with the same content, and (ii) *X* and *Y* having the very same belief. Why does Schechtman say that only the second is relevant to personal identity? Do you agree? What if *X*, at one time, had *all* the same beliefs, in sense (i), as person *Y* at some later time – would *that* be irrelevant to whether *X* and *Y* are the same person?

We now come to the notion of a 'quasi memory' or *q*-memory. It is this notion in particular that Schechtman will be attacking in detail later on. The original circularity objection, remember, is that in order for a mental state *m* to be a genuine memory of some past experience *e*, the person in state *m* must be the very same person as the one who originally had the experience *e*. Schechtman says at [d]→ that 'identity theorists' (by which she means 'those who think that personal identity can be defined in terms of psychological continuity') accept that this is true of memories, but claim that psychological continuity can be defined instead in terms of *q*-memories; and it is *not* the case that in order for person *X*'s mental state *m* to be a *q*-memory of experience *e*, *X* must be identical with – must be the same person as – the person who originally had experience *e*.

3. Using your own example, state in your own words the distinction identity theorists make between memory, delusion and *q*-memory (see [e]→).
4. (i) Describe in your own words the objection to which Parfit is responding at [f]→, and his response. (ii) Schechtman says at [g]→ that she will argue, against Parfit, that 'simply deleting the ''nametag'' from a memory is not sufficient to make it nondelusional'. Explain in your own words why this claim contradicts Parfit's view.

Section III

Schechtman's argument against identity theories takes the form of a dilemma, clearly stated in the paragraph at [h]→: either Jane's *q*-memory will have only the same *visual* content as Casey's memory, or it will be qualitatively *exactly* similar to Casey's. Schechtman says: 'Neither of these alternatives will allow quasi memory to do the work it is supposed to do.'

5. Describe in your own words why, according to Schechtman, neither alternative will 'allow quasi memory to do the work it is supposed to do'.

Schechtman's argument, then, attempts to establish the claim that 'facts about whose memory a given memory is are an integral part of its qualitative content' [i]→. We cannot, even in principle – contrary to what Parfit claims – 'separate off' the *content* of a memory from the fact about whose memory it is. If we think of Jane's *q*-memory as just a flow of purely visual images, then 'it seems arbitrary to call it even the same apparent memory' as Casey's: what Jane experiences when *q*-remembering the trip to the cinema is qualitatively very different to what Casey experiences when he remembers the trip. But if we think of Jane's *q*-memory of the trip as qualitatively the same in *every* respect as Casey's experience when he remember the trip, then Jane's experience is delusional: it will include thoughts such as 'these are *my* children', and '*I* have been to this restaurant before', and so on, all of which are, in Jane's case, false.

6. Does Schechtman's argument establish that *all* (quasi) memories have this feature, or only that some do? Does she need the former claim, or only the latter, in order to establish that psychological continuity cannot be non-circularly defined?

Section IV

Finally, Schechtman attempts to diagnose what has gone wrong with psychological-continuity theories of personal identity, and sketches an outline of a positive view which does not, Schechtman claims, make the same mistake.

The notion of a 'diagnosis' in a philosophical context is a useful one. If someone wakes up with a fever and blotches on his skin, he knows there is *something* wrong with him, but he might not know what it is. When the doctor diagnoses him, she makes a claim about what the underlying cause of the symptoms is, and (if the patient is lucky) will therefore be able to cure him, by prescribing antibiotics or whatever.

A counter-example to a philosophical theory can be analogous to fever and blotches on the skin: it tells us *that* there is something wrong with the theory, and hence that the theory cannot be true as it stands. But it might not, just by itself – or at least, not in an obvious way – tell us what the underlying 'cause' of the problem is,

and it might therefore be unclear whether there is a fairly easy way to fix the theory, or how one might go about constructing a new theory that avoids the problem. A diagnosis of the problem explains to us what its underlying 'cause' is, and therefore helps us to see how to respond to it.

This is precisely what Schechtman does in the four paragraphs from [j]→. She explains not only why we cannot expect to be able to cash psychological continuity out in a non-circular way, but also why it *seems* right to think, mistakenly, that personal identity *does* depend on psychological continuity. Roughly, she distinguishes between what she calls the 'reidentification question' and the 'self-knowledge question' (this is a distinction first introduced in the opening paragraph of the paper). She argues that philosophers want to provide a non-circular criterion of personal identity because they want to answer the re-identification question; but the intuitions that support thinking of personal identity as a matter of psychological rather than bodily continuity come from thinking about the self-knowledge question. So to attempt to provide a non-circular criterion of personal identity in terms of psychological continuity is in effect to try to provide a single answer to two very different questions.

7. State in your own words what, according to Schechtman, the connection is (i) between the 'self-knowledge question' and 'our view of persons as subjects', and (ii) between the 'reidentification question' and viewing persons as 'objects'.

Let's return to the issue of reductionism. Schechtman's own view seems to be this. A *reductionist* criterion of personal identity can be given, where we think of persons as 'objects'; and that criterion will be 'bodily' rather than psychological (see [k]→). (Notice, however, that she does not elaborate on what she means by a 'bodily criterion'; nor does she attempt to argue that such a criterion can in fact be given.) However, were we to think of persons as 'subjects', a psychological criterion of personal identity can be given, but it will *not* provide a 'reductionist analysis of persons' (see [l]→).

8. (i) State in your own words the central features of the positive view of personal identity which Schechtman sketches in the last few pages of the paper. (ii) In *Reasons and Persons*, Parfit claims: 'On the Non-Reductionist View, a person is a separately existing entity, distinct from his brain and body, and his experiences' (Parfit 1984: 275). Do you think Schechtman would or needs to agree with this claim? Why or why not?

Finally, it's worth thinking about Schechtman's paper in the context of Parfit's overall point, that survival is more important than personal identity. Schechtman ignores survival – as opposed to identity – altogether. The conclusion of her negative argument is that facts about psychological continuity presuppose facts about personal *identity*, and her positive claim is, again, a claim about personal identity. So, if one is convinced by Parfit's paper that we ought to abandon the language of identity, one *might* conclude that Schechtman's paper is just irrelevant.

9. Do you think that (i) Schechtman's negative argument and (ii) her positive proposal could be recast in a way that replaces talk of identity with talk of survival? Why or why not? What consequences does your answer have, in the light of Parfit's discussion of the divided-brain case and one or more of the other thought-experiments he discusses?

Further Reading

A good place to start might be chapter 6 of Guttenplan, Hornsby and Janaway's *Reading Philosophy* (2003). This contains abridged versions of Locke's classic discussion of personal identity (1690) and Williams 1970, along with commentaries and questions for you to work through. It's well worth reading the Locke in full as well though – as are the classic responses to Locke by Thomas Reid (1785) and Joseph Butler (1736).

Parfit has a lot more to say about personal identity and survival in his *Reasons and Persons* (1984), chapters 10–15. One area which both of the texts reproduced above touch on but do not explore in any detail is the 'bodily' criterion of personal identity: a criterion that Parfit rejects but which, as applied to persons considered as 'objects', Schechtman endorses. For more on the bodily criterion, see for example Johnston 1987 and Nagel 1986, chapter 3. Schechtman's positive view connects in interesting ways with Nagel's; again, see his 1986, chapter 3 (and also chapters 1 and 2).

An excellent edited collection on personal identity is Perry 1975. It contains some of the papers mentioned here (Locke, Butler, Reid, Williams), together with several others and a useful introduction. A good textbook is Noonan 2003; you might also try Hanley 1997, chapters 4 and 5 – particularly if you are a fan of *Star Trek*.

Essay Questions

1. Is survival more important than personal identity?
2. Can the circularity objection to psychological-continuity theories of personal identity be met?
3. Would we be better off if we abandoned or restricted the use of the language of personal identity in favour of different ways of thinking?
4. Can we be morally responsible for actions we cannot remember performing? What consequences, if any, does your answer have for personal identity?
5. What form might a 'bodily continuity' or 'bodily identity' theory of personal identity take? Could bodily continuity or identity be a necessary condition of personal identity? (That is, could it be the case that personal identity *requires* bodily continuity or bodily identity?) Could it be a sufficient condition? (That is, could it be the case that bodily continuity or identity *guarantees* personal identity?)