Derek Parfit and Godfrey Vesey

BRAIN TRANSPLANTS AND PERSONAL IDENTITY
A Dialogue

Derek Parfit is an English philosopher who was educated and now teaches at Oxford University. He has made outstanding contributions to the subjects of ethical theory and the problem of personal identity. His major work is *Reasons and Persons* (1984). Godfrey Vesey was educated at Cambridge University and is a professor of philosophy at Open University. His principal works are *Perception* (1971) and *Personal Identity* (1974).

In this dialogue, Vesey introduces the problem of split-brain transplants. That is, a brain is divided into two, and half is put into each of two other people's brainless heads. Does the original person survive? Parfit then responds by developing his ideas of personal identity as psychological identity.

BRAIN TRANSPLANTS

In 1973 in the *Sunday Times* there was a report of how a team from the Metropolitan Hospital in Cleveland under Dr. R. J. White had successfully transplanted a monkey's head on to another monkey's body.[1] Dr. White was reported as having said, 'Technically a human head transplant is possible', and as hoping that 'it may be possible eventually to transplant *parts* of the brain or other organs inside the head.'

The possibility of brain transplants gives rise to a fascinating philosophical problem. Imagine the following situation:

> Two men, a Mr Brown and a Mr Robinson, had been operated on for brain tumours and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's body and Brown's brain, eventually regains consciousness. Let us call the latter 'Brownson'. Upon regaining consciousness Brownson exhibits great shock and surprise at the appearance of his body. Then, upon seeing Brown's body, he exclaims incredulously 'That's me lying there!' Pointing to himself he says 'This isn't my body; the one over there is!' When asked his name he automatically replies 'Brown'. He recognizes Brown's wife and family (whom Robinson had never met), and is able to describe in detail events in Brown's life, always describing them as events in his own life. Of Robinson's past life he evinces no knowledge at all. Over a period of time he is observed to display all of the personality traits, mannerisms, interests, likes and dislikes, and so on, that had previously characterized Brown, and to act and talk in ways completely alien to the old Robinson.[2]

Reprinted by permission of Open University Press, from "Brain Transplants and Personal Identity," in *Philosophy in the Open*, ed. Godfrey Vesey, 1974.

The next step is to suppose that Brown's brain is not simply transplanted whole into someone else's brainless head, but is divided in two and half put into each of *two* other people's brainless heads. The same memory having been coded in many parts of the cortex, they *both* then say they are Brown, are able to describe events in Brown's life as if they are events in their own lives, etc. What should we say now?

The implications of this case for what we should say about personal identity are considered by Derek Parfit in a paper entitled 'Personal Identity'. Parfit's own view is expressed in terms of a relationship he calls 'psychological continuity'. He analyses this relationship partly in terms of what he calls '*q*-memory' ('*q*' stands for 'quasi'). He sketches a definition of '*q*-memory' as follows:

> 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.[3]

The significance of this definition of *q*-memory is that *two* people can, in theory, *q*-remember doing what only one person did. So two people can, in theory, be psychologically continuous with one person.

Parfit's thesis is that there is nothing more to personal identity than this 'psychological continuity'. This is *not* to say that whenever there is a sufficient degree of psychological continuity there is personal identity, for psychological continuity could be a one-two, or 'branching', relationship, and we are able to speak of 'identity' only when there is a one-one relationship. It *is* to say that a common belief—in the special nature of personal identity—is mistaken.

In the discussion that follows I began by asking Parfit what he thinks of this common belief.

PERSONAL IDENTITY

Vesey: Derek, can we begin with the belief that you claim most of us have about personal identity? It's this: whatever happens between now and some future time either I shall still exist or I shan't. And any future experience will either be my experience or it won't. In other words, personal identity is an all or nothing matter: either I survive or I don't. Now what do you want to say about that?

Parfit It seems to me just false. I think the true view is that we can easily describe and imagine large numbers of cases in which the question, 'Will that future person be me—or someone else?', is both a question which doesn't have any answer at all, and there's no puzzle that there's no answer.

Vesey: Will you describe one such case.

Parfit One of them is the case discussed in the correspondence material, the case of division in which we suppose that each half of my brain is to be transplanted into a new body and the two resulting people will both seem to remember the whole of my life, have my character and be psychologically continuous with me in every way. Now in this case of division there were only three possible answers to the question, 'What's going to happen to *me*?' And all three of them seem to me open to very serious objections. So the conclusion to be drawn from the case is that the question of what's going to happen to me, just doesn't have an answer. I think the case also shows that that's not mysterious at all.

Vesey: Right, let's deal with these three possibilities in turn.

Parfit Well, the first is that I'm going to be both of the resulting people. What's wrong with that answer is that it leads very quickly to a contradiction.

Vesey: How?

Parfit The two resulting people are going to be different people from each other. They're going to live completely different lives. They're going to be as different as any two people are. But if they're different people from each other it can't be the case that I'm going to be both of them. Because if I'm both of them, then one of the resulting people is going to be the same person as the other.

Vesey: Yes. They can't be different people and be the same person, namely me.

Parfit Exactly. So the first answer leads to a contradiction.

Vesey: Yes. And the second?

Parfit Well, the second possible answer is that I'm not going to be both of them but just one of them. This

doesn't lead to a contradiction, it's just wildly implausible. It's implausible because my relation to each of the resulting people is exactly similar.

Vesey: Yes, so there's no reason to say that I'm one rather than the other?

Parfit It just seems absurd to suppose that, when you've got exactly the same relation, one of them is identity and the other is nothing at all.

Vesey: It does seem absurd, but there are philosophers who would say that sort of thing. Let's go on to the third.

Parfit Well, the only remaining answer, if I'm not going to be both of them or only one of them, is that I'm going to be neither of them. What's wrong with this answer is that it's grossly misleading.

Vesey: Why?

Parfit If I'm going to be neither of them, then there's not going to be anyone in the world after the operation who's going to be me. And that implies, given the way we now think, that the operation is as bad as death. Because if there's going to be no one who's going to be me, then I cease to exist. But it's obvious on reflection that the operation isn't as bad as death. It isn't bad in any way at all. That this is obvious can be shown by supposing that when they do the operation only one of the transplants succeeds and only one of the resulting people ever comes to consciousness again.

Vesey: Then I think we would say that this person is me. I mean we'd have no reason to say that he wasn't.

Parfit On reflection I'm sure we would all think that I would survive as that one person.

Vesey: Yes.

Parfit Yes. Well, if we now go back to the case where both operations succeed . . .

Vesey: Where there's a double success . . .

Parfit It's clearly absurd to suppose that a double success is a failure.

Vesey: Yes.

Parfit So the conclusion that I would draw from this case is firstly, that to the question, 'What's going to happen to me?', there's no true answer.

Vesey: Yes.

Parfit Secondly, that if we decide to say one of the three possible answers, what we say is going to obscure the true nature of the case.

Vesey: Yes.

Parfit And, thirdly, the case isn't in any way puzzling. And the reason for that is this. My relation to each of the resulting people is the relation of full psychological continuity. When I'm psychologically continuous with only one person, we call it identity. But if I'm psychologically continuous with two future people, we can't call it identity. It's not puzzling because we know exactly what's going to happen.

Vesey: Yes, could I see if I've got this straight? Where there is psychological continuity in a one-one case, this is the sort of case which we'd ordinarily talk of in terms of a person having survived the operation, or something like that.

Parfit Yes.

Vesey: Now what about when there is what you call psychological continuity—that's to say, where the people seem to remember having been me and so on—in a one-two case? Is this survival or not?

Parfit Well, I think it's just as good as survival, but the block we have to get over is that we can't say that anyone in the world after the operation is going to be me.

Vesey: No.

Parfit Well, we can say it but it's very implausible. And we're inclined to think that if there's not going to be anyone who is me tomorrow, then I don't survive. What we need to realize is that my relation to each of those two people is just as good as survival. Nothing is missing at all in my relation to both of them, as compared with my relation to myself tomorrow.

Vesey: Yes.

Parfit So here we've got survival without identity. And that only seems puzzling if we think that identity is a further fact over and above psychological continuity.

Vesey: It is very hard not to think of identity being a further fact, isn't it?

Parfit Yes, I think it is. I think that the only way to get rid of our temptation to believe this is to consider many more cases than this one case of division. Perhaps I should give you another one. Suppose that the following is going to happen to me. When I die in a normal way, scientists are going to map the states of all the cells in my brain and body and after a few

months they will have constructed a perfect duplicate of me out of organic matter. And this duplicate will wake up fully psychologically continuous with me, seeming to remember my life with my character, etc.

Vesey: Yes.

Parfit Now in this case, which is a secular version of the Resurrection, we're very inclined to think that the following question arises and is very real and very important. The question is, 'Will that person who wakes up in three months be me or will he be some quite other person who's merely artificially made to be exactly like me?'

Vesey: It does seem to be a real question. I mean in the one case, if it is going to be me, then I have expectations and so on, and in the other case, where it isn't me, I don't.

Parfit I agree, it seems as if there couldn't be a bigger difference between it being me and it being someone else.

Vesey: But you want to say that the two possibilities are in fact the same?

Parfit I want to say that those two descriptions, 'It's going to be me' and 'It's going to be someone who is merely exactly like me', don't describe different outcomes, different courses of events, only one of which can happen. They are two ways of describing one and the same course of events. What I mean by that perhaps could be shown if we take an exactly comparable case involving not a person but something about which I think we're not inclined to have a false view.

Vesey: Yes.

Parfit Something like a club. Suppose there's some club in the nineteenth century . . .

Vesey: The Sherlock Holmes Club or something like that?

Parfit Yes, perhaps. And after several years of meeting it ceases to meet. The club dies.

Vesey: Right.

Parfit And then two of its members, let's say, have emigrated to America, and after about fifteen years they get together and they start up a club. It has exactly the same rules, completely new membership except for the first two people, and they give it the same name. Now suppose someone came along and said: 'There's a real mystery here, because the following question is one that must have an answer. But how can we answer it?' The question is, 'Have they started up the very same club—is it the same club as the one they belonged to in England—or is it a completely new club that's just exactly similar?'

Vesey: Yes.

Parfit Well, in that case we all think that this man's remark is absurd; there's no difference at all. Now that's my model for the true view about the case where they make a duplicate of me. It seems that there's all the difference in the world between its being me and its being this other person who's exactly like me. But if we think there's no difference at all in the case of the clubs, why do we think there's a difference in the case of personal identity, and how can we defend the view that there's a difference?

Vesey: I can see how some people would defend it. I mean, a dualist would defend it in terms of a soul being a simple thing, but . . .

Parfit Let me try another case which I think helps to ease us out of this belief we're very strongly inclined to hold.

Vesey: Go on.

Parfit Well, this isn't a single case, this is a whole range of cases. A whole smooth spectrum of different cases which are all very similar to the next one in the range. At the start of this range of cases you suppose that the scientists are going to replace one per cent of the cells in your brain and body with exact duplicates.

Vesey: Yes.

Parfit Now if that were to be done, no one has any doubt that you'd survive. I think that's obvious because after all you can *lose* one per cent of the cells and survive. As we get further along the range they replace a larger and larger percentage of cells with exact duplicates, and of course at the far end of the range, where they replace a hundred per cent, then we've got my case where they just make a duplicate out of wholly fresh matter.

Vesey: Yes.

Parfit Now on the view that there's all the difference in the world between its being me and its being this other person who is exactly like me, we ought in consistency to think that in some case in the middle of that range, where, say, they're going to replace fifty per cent, the same question arises: it is going to be me

or this completely different character? I think that even the most convinced dualist who believes in the soul is going to find this range of cases very embarrassing, because he seems committed to the view that there's some crucial percentage up to which it's going to be him and after which it suddenly ceases to be him. But I find that wholly unbelievable.

Vesey: Yes. He's going to have to invent some sort of theory about the relation of mind and body to get round this one. I'm not quite sure how he would do it. Derek, could we go on to a related question? Suppose that I accepted what you said, that is, that there isn't anything more to identity than what you call psychological continuity in a one-one case. Suppose I accept that, then I would want to go on and ask you, well, what's the philosophical importance of this?

Parfit The philosophical importance is, I think, that psychological continuity is obviously, when we think about it, a matter of degree. So long as we think that identity is a further fact, one of the things we're inclined to think is that it's all or nothing, as you said earlier. Well, if we give up that belief and if we realize that what matters in my continued existence is a matter of degree, then this does make a difference in actual cases. All the cases that I've considered so far are of course bizarre science fiction cases. But I think that in actual life it's obvious on reflection that, to give an example, the relations between me now and me next year are much closer in every way than the relations between me now and me in twenty years. And the sorts of relations that I'm thinking of are relations of memory, character, ambition, intention—all of those. Next year I shall remember much more of this year than I will in twenty years. I shall have a much more similar character. I shall be carrying out more of the same plans, ambitions and, if that is so, I think there are various plausible implications for our moral beliefs and various possible effects on our emotions.

Vesey: For our moral beliefs? What have you in mind?

Parfit Let's take one very simple example. On the view which I'm sketching it seems to me much more plausible to claim that people deserve much less punishment, or even perhaps no punishment, for what they did many years ago as compared with what they did very recently. Plausible because the relations between them now and them many years ago when they committed the crime are so much weaker.

Vesey: But they are still the people who are responsible for the crime.

Parfit I think you say that because even if they've changed in many ways, after all it was just as much they who committed the crime. I think that's true, but on the view for which I'm arguing, we would come to think that it's a completely trivial truth. It's like the following truth: it's like the truth that all of my relatives are just as much my relatives. Suppose I in my will left more money to my close relatives and less to my distant relatives; a mere pittance to my second cousin twenty-nine times removed. If you said, 'But that's clearly unreasonable because all of your relatives are just as much your relatives', there's a sense in which that's true but it's obviously too trivial to make my will an unreasonable will. And that's because what's involved in kinship is a matter of degree.

Vesey: Yes.

Parfit Now, if we think that what's involved in its being the same person now as the person who committed the crime is a matter of degree, then the truth that it was just as much him who committed the crime will seem to us trivial in the way in which the truth that all my relatives are equally my relatives is trivial.

Vesey: Yes. So you think that I should regard myself in twenty years' time as like a fairly distant relative of myself?

Parfit Well, I don't want to exaggerate; I think the connections are much closer.

Vesey: Suppose I said that this point about psychological continuity being a matter of degree—suppose I said that this isn't anything that anybody denies?

Parfit I don't think anybody does on reflection deny that psychological continuity is a matter of degree. But I think what they may deny, and I think what may make a difference to their view, if they come over to the view for which I'm arguing—what they may deny is that psychological continuity is all there is to identity. Because what I'm arguing against is this further belief which I think we're all inclined to hold even if we don't realize it. The belief that however much we change, there's a profound sense in

which the changed us is going to be just as much us. That even if some magic wand turned me into a completely different sort of person—a prince with totally different character, mental powers—it would be just as much me. That's what I'm denying.

Vesey: Yes. This is the belief which I began by stating, and I think that if we did lose that belief that would be a change indeed.

NOTES

1. *Sunday Times*, 9 December, 1973, p. 13.
2. Shoemaker (1963), pp. 23–24.
3. Parfit (1971), p. 15.

FOR FURTHER REFLECTION

1. Is Parfit's analysis of the problem of the split brain case plausible? Do you think that he correctly construes the issue, or does he already presuppose controversial premises? Explain
2. What are the three possible answers to the question "What's going to happen to me?" in the split-brain case, and how does Parfit treat them? What is Parfit's solution to the problem? Could you suggest another solution to the problem of personal identity?