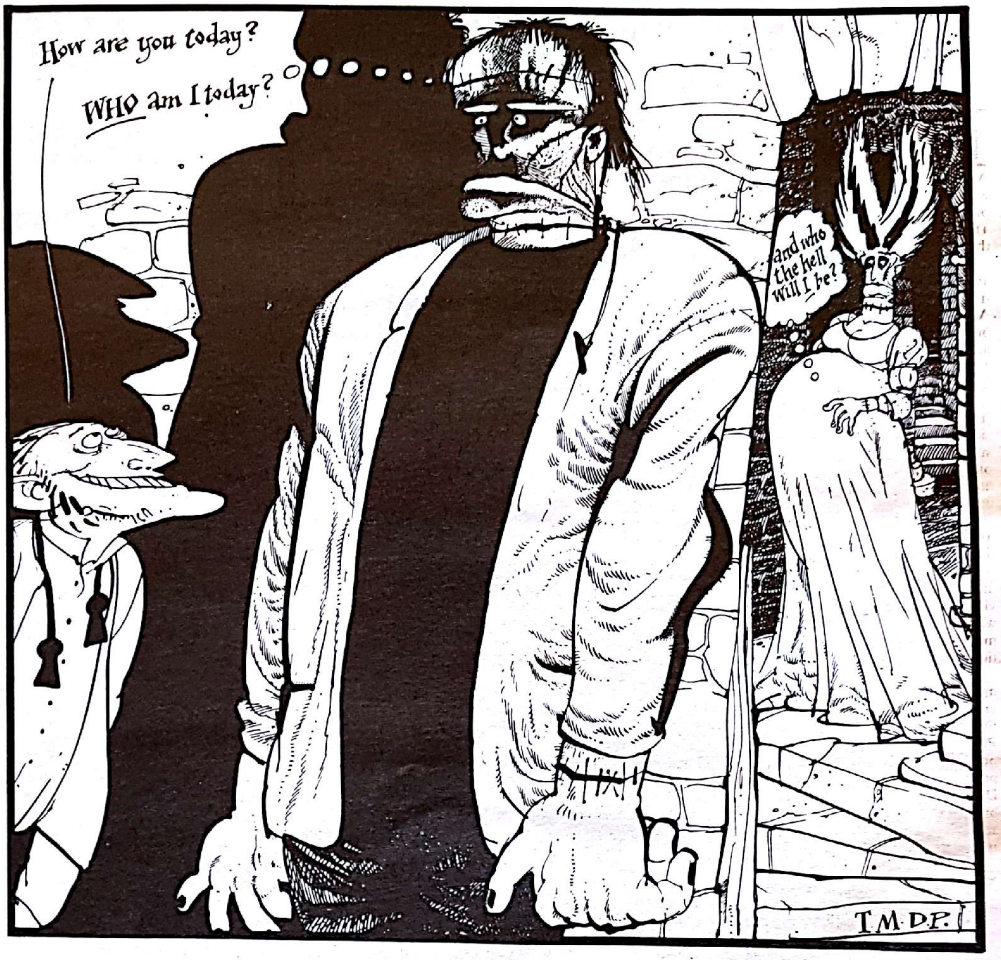
Who do you think you are?

The second of a series from the Herbert Spencer Lectures at Oxford University

Derek Parfit on the unimportance of personal identity



Most of us are specially concerned about our own future. What, if anything, gives us our reasons for such concern?

We can explain this question with an imaginary case. Suppose that some strange person, while concerned about his whole future, cared especially about any pain that he would feel on a future Tuesday. That would be irrational. The fact that he would feel a pain on *Tuesday* is no reason to care about it more. But what about the fact the *he* would feel this pain? Is *that* a reason to care about it more?

Most of us would answer Yes. We care about our future because it will be *our* future. Though this is the natural view, I believe it to be mistaken. On my view, we should care about our future, not because it will be ours, but for other reasons. Personal identity is not what matters.

In the first part of this talk, I have sketched a *reductionist* account of what personal identity is. On this account, our continued existence over time just consists in certain other facts. I shall now argue that, on a reductionist account, personal identity cannot be what matters. My argument assumes that, when one fact just consists in certain others, it can only be these other facts which matter. This assumption we can call *reductionism about significance*.

It may help to illustrate this assumption in a slightly different area. Consider people who, because of brain damage, are irreversibly unconscious, but whose hearts are still beating, and whose other organs are still functioning. It has become important to decide how we ought to treat such people.

Suppose we believe that such people are still alive. This fact could be understood in a reductionist way. It may not be the same as the fact that these people's hearts are still beating, and their other organs are still functioning. But it is not an independent or separately obtaining fact. Such a person's being still alive, though irreversibly unconscious, just consists in these other facts.

We can now appeal to reductionism about significance. Since this fact just consists in these other facts, it can only be these other facts which matter. We should ask: If people are irreversibly unconscious, does it matter in itself that their hearts and other organs are still functioning? Is that either good for these people, or morally significant? If we answer No, we should conclude that it cannot matter that these people are still alive. We would be right to stop their hearts from beating, if that is what their relatives prefer.

Some of us would take a different view. If such people can be correctly called "alive", stopping their hearts would be killing them. And we may believe that, except in self-defence, it is always wrong to kill. We may then conclude that we should help these hearts to go on beating, perhaps for many years.

On reductionism about significance, that is irrational. If it has no significance, in itself, that someone's heart is beating, this fact cannot be significant when and because it constitutes the fact that this person is still alive. The beating of a heart can only matter when and because it makes possible some independent fact which matters, such as a person's being conscious.

Return now to personal identity. I believe that, though we have good reason to care about our future, that is not because it is *our* future. That can best be seen in certain imaginary cases.

Suppose that because of damage to my heart and lungs, I am tied to a heart-lung machine. A new possibility arises. My twin brother develops a fatal brain disease. Surgeons tell me that, when my brother's brain ceases to function, my head could be separated from the rest of my body, and be grafted on to my brother's headless body. The resulting person, they point out, would be freed from the heart-lung machine. After getting my brother's consent, I decide to accept this operation.

According to some philosophers, I would be making a terrible mistake. On their view, a person continues to exist if and only if that person's body continues to exist. These philosophers claim that, if my head were grafted on to my brother's body, it would be my brother who would wake up, falsely believing that he was me. Other philosophers claim that it would be risky to accept this operation, since the person who woke up might not be me. A third group warn me that, in such a case, there would be no answer to the question who the person waking up would be.

That question, I believe, has no importance. If this operation were performed, the person with my head tomorrow would not only believe that he was me, seem to remember living my life, and be in every other way psychologically just like me. These facts would also have their normal cause, since this person would have my brain. And his life would be very like the life which I would have lived, if my heart and lungs had been repaired. I believe that, given these facts, I should accept this operation. It is irrelevant whether this person would be me.

It may be objected: "If he would not be you, you would have ceased to exist. Surely that matters!" But, if this person would not be me, that fact would just consist in another fact. It would just consist in the fact that my head will have been connected to the rest of my brother's body. Is that second fact, on its own, important? Can it matter in itself that the blood that will keep my brain alive will circulate, not through a heart-lung machine, but through my brother's heart and lungs? Can it matter in itself that my head will be connected, not to my own body, but to another body that is just like mine? I believe that, in themselves, these facts cannot rationally be thought to matter. Or, at least, they cannot matter much. They would be far outweighed by the fact that, once my head has been connected to my brother's body, the resulting person would be freed from the heart-lung machine.

One way to test that claim would be this. Suppose we believed that this person *would* be me. Would we then believe that it would matter greatly that my head would have been connected to a new body? We would not. We would regard my receiving a new torso, and new limbs, as like any lesser transplant, such as receiving a new heart, or new kidneys. If it would matter greatly that what will be replaced is not just one or two organs, but my whole body below the neck, that could only be because, if that happened, the resulting person would *not* be me.

Since that is so, if we accept reductionism about significance, we should conclude that neither of these facts could matter greatly. The fact that this person wouldn't be me, though not the same as the fact that my head would have been given a new body, wouldn't be a further difference in what happens. Since it wouldn't in itself matter that my head would be given a new body, and that would be all there was to the fact that the resulting person wouldn't be me, this fact wouldn't matter either.

That conclusion can be supported in a different way. If one fact just consists in certain others, the first fact is, in a sense, conceptual. Thus, if someone is irreversibly unconscious, but his heart is still beating, it is a conceptual fact that this person is still alive. And if my head were given a new body, it would be a conceptual fact (if it were a fact at all) that the resulting person would not be me. In calling these facts conceptual, I don't mean that they are facts about our concepts. That this unconscious person is alive is a fact about this person. But, if we have already claimed that this person's heart is beating, our claim that he is alive gives no further information about reality. It only gives further information about our use of the words "person" and "alive".

Suppose we agree that it does not matter, in itself, that such a person's heart is still beating. Could we claim that, in another way, this does matter, because it makes it correct to say that this person is still alive? And suppose we agree that it would not matter, in itself, if my head were given a new body. Could this matter if, and because, it makes it correct to say that the resulting person would not be me?

If we answer Yes, we are treating language as more important than reality. We are claiming that, even if some fact does not in itself matter, it may matter if and because it allows a certain word to be applied. On such a view, it always matters whether a person can be correctly called "alive", or called "not the same" as some previously existing person. It is irrelevant if this person's being alive, or not the same person, just consist in other facts which are in themselves trivial. What matters is *that* these words apply, not *why* they apply.

This view seems to me irrational. What matters is reality, not how it is described. In deciding whether my head should be given a new body, I should try to predict the physical and mental state of the resulting person. But I need not ask whether that person could be correctly called me. That would not be a difference in what happens.

That may seem a false distinction. What matters, we might say, is whether the resulting person would *be* me. But that person would be me if and only if he could be correctly called me. So, in asking what he could be called, we are not merely asking a conceptual question. We *are* asking about reality.

This objection fails to distinguish two kinds of case. Suppose that I ask my dentist whether some treatment will be painful. That is a factual question. Since pain can be called "pain", I *could* ask my question in a different way. I could say, "While I am being treated, will it be correct to describe me as in pain?" But that would be misleading. It would suggest that I am asking how we use the word "pain".

In a different case, I might ask that conceptual question. Suppose I know that, when I cross the Channel, I shall feel sea-sick, as I always do. I might wonder whether that sensation could be correctly called "pain". Here too, I could ask my question in a different way. I could say, "While I cross the Channel, shall I be in pain?" But that would be misleading, since it would suggest that I asking what will happen.

In the dental case, I don't know what conscious state I shall be in. There are different possibilities. In the Channel crossing case, there aren't different possibilities. I already know what state I shall be in. I am merely asking whether that state could be redescribed in a certain way.

It matters whether, while receiving the dental treatment, I shall be in pain. And it matters whether, while crossing the Channel, I shall feel sea-sick. But it doesn't matter whether, in feeling sea-sick, I can be said to be in pain.

Return now to our main example. Suppose I have accepted our imagined operation, and I wonder whether the resulting person will be me. Is this like the case of the treatment, or the Channel crossing? Am I asking what will happen, or whether some concept can be applied?

On my view, I am asking the second. I already know what's going to happen. There will be someone with my head and my brother's body. It's merely a verbal question whether that person will be me. And that's why, even if he won't be me, that doesn't matter.

Continued on page 20

THE TIMES HIGHER DECEMBER 11 1992

Who do you think you are?

Continued from page 19

It may still be hard to believe that this is a verbal question. If I don't know whether I shall exist tomorrow, it may be hard to believe that I know what's going to happen. But what is it that I don't know? If there are different possibilities, in what would the difference consist? If I had a soul, or Cartesian Ego, there might be different possibilities. Perhaps when my head is grafted on to my brother's body, my soul would keep its intimate relation with my brain, or perhaps my brother's soul would take over. But, I believe, there are no such entities. What else could the difference be? When the person with my head wakes up tomorrow, what could make it true, or false, that he is me?

It may be said that, in asking what will happen, I am asking what I can expect. Can I expect to wake up again tomorrow? If that person will be in pain, can I expect to feel that pain? But this does not help. These are just other ways of asking whether that person will or will not be me. In appealing to what I can expect, we do not explain what would make these different possibilities.

It is natural to reply that this difference doesn't need to be explained. We may think it enough to say: Perhaps I shall exist tomorrow, and perhaps I shan't. It may seem that these must be different possibilities.

That, however, is an illusion. If I shall exist tomorrow, that fact must consist in certain others. For there to be two possibilities, so that it might be either true or false that I shall exist tomorrow, there must be some other difference between these possibilities. There would be such a difference, for example, if my brain and body might either survive the night, or be blown to pieces. But, in our imagined case, there is no such other difference. The full facts are these. After I am made unconscious, my head will be grafted on to my brother's body, and the person who wakes up will be psychologically just like me. There is no further essence of me, or property of me-ness, which either might or might not be there.

If we turn to the conceptual level, there *are* different possibilities. Perhaps that future person could be correctly called me. Perhaps he could be correctly called my brother. Or perhaps neither would be correct. That, however, is the only way in which it could be either true, or false, that this person would be me.

The illusion may persist. Even when I know the other facts, I may want reality to go in one of two ways. I may want it to be true that I shall exist tomorrow. But all that could be true is that we use language in one of two ways. Can it be rational to care about that?

I turn now to a different objection. I have argued that, in deciding whether to accept this operation, I need not ask whether the resulting person would be me. In such a case, personal identity does not matter. It may be said: "By choosing this example, you are cheating. Perhaps you should accept this operation. But that is because the resulting person *would* be you. So this case cannot show that identity is not what matters."

I accept part of this objection. Animals continue to exist if there continues to exist, and to function, the most important parts of their bodies. It would be reasonable to claim that, at least in the case of human beings, the brain is so important that its survival would be the survival of this human being. On this view, in my imagined case, the person with my head tomorrow would be me. And this is what, on reflection, most of us would believe.

My own view is slightly different. I would state this view, not as a claim about reality, but as a conceptual claim. On my view, it would not be incorrect to call this person me; and this would be the best description of this case.

Even if this person can be claimed to be me, I would still argue that this is not what matters. What is important is not identity, but one or more of the other relations in which identity consists. But, when identity coincides with those other relations, it is harder to decide whether we accept that argument's conclusion. We need a case where, though some future person would *not* be me, this would not matter.

We are now assuming that, for me to survive, it is enough that my brain survives, and continues to function. On this assumption, the argument can be transferred to another case. And this will make it, in one way, stronger.

We can first note that, just as I could survive with less than my whole body, I could survive with less than my whole brain. People have survived, and with little change, even when, through a stroke or injury, they have lost the use of half their brain. Let us next suppose that each half of my brain could support an ordinary mental life. (Since we are testing our beliefs about what matters, it is irrelevant whether this is actually true.)

We can now compare two possible operations. In the first, after half my brain is destroyed, the other half would be successfully transplanted into the empty skull of a body that is just like mine. Given our assumptions, we should conclude that, here too, I would survive. Since I would survive if my brain were transplanted, and I would survive with only half my brain, I would survive if that remaining half were transplanted. So, in this *One-Sided Case*, the resulting person would be me.

Consider next the *Two-Sided Case*, or My Division. Both halves of my brain would be successfully transplanted, into different bodies. Two people would wake up, each of whom has half my brain, and is, both physically and psychologically, just like me.

Since these would be two different people, it can't be true that each of them is me. That would be a contradiction. If each of them was me, each would be one and the same person: me. So they couldn't be two different people.

Could it be true that only one of them is me? That is not a contradiction. But, since I have the same relation to each of these people, there is nothing that could make me one of them rather than the other. It cannot be true, of either of these people, that he is the one who could be correctly called me.

How should I regard these two operations? Would they preserve what matters in survival? In the One-Sided Case, the one resulting person would be me. My relation to that person is just an instance of the relation between me now and myself tomorrow. That relation would contain what matters. In the Two-Sided Case, my relation to that person would be just the same. So this relation must still contain what matters. Nothing is missing. But that person cannot here be claimed to be me. So identity cannot be what matters.

We may object that, if that person isn't me, something *is* missing. *I'm* missing. That may seem to make all the difference. How can everything still be there if *I'm* not there?

Everything is still there. The fact that I'm not there is not a real absence. The relation between me and that future person is in itself the same. As in the One-Sided Case, he has half my brain, and he is just like me. The difference is only that, in this Two-Sided Case, I also have the same relation to the other resulting person. Why am I not there? The explanation is only this. When this relation holds between me now and a single person in the future, we can be called one and the same person. When this relation holds between me now and *two* future people, I cannot be called one and the same as each of these people. But that is not a difference in the nature or the content of this relation. In the One-Sided Case, where half my brain will be successfully transplanted, my prospect is survival. That prospect contains what matters. In the Two-Sided Case, where both halves will be successfully transplanted, nothing would be lost.

It can be hard to believe that identity is not what matters. But that is easier to accept when we see why, in this example, it is true. It may help to consider this analogy. Imagine a community of persons who resemble us, but with two exceptions. First, because of facts about their reproductive system, each couple has only two children, who are always twins. Second, because of special features of their psychology, it is of great importance for the development of each child that it should not, through the death of its sibling, become an only child. Such children suffer severe psychological damage. It is thus believed, in this community, that it matters greatly that each child should have a twin.

Now suppose that, because of some biological change, some of the children in this community are born as triplets. Should their parents think this a disaster, because these children don't have twins? Clearly not. Why do these children not have twins? Only because they each have *two* siblings. Since each child has two siblings, the trio must be called, not twins, but triplets. But none of these children suffer damage as an only child. So these people should revise their view. What matters isn't having a twin: it's having at least one sibling.

In our view about identity, we should make a similar revision. What matters isn't that there will be someone alive who will be me. It is rather that there will be at least one living person who will be psychologically continuous with me as I am now, and/or who will have enough of my brain. When there will be only one such person, he can be described as me. When there will be two such people, we cannot claim that each will be me. But that is as trivial as the fact that, if I had two identical siblings, they couldn't be called my twins.

If personal identity isn't what matters, we must ask what does matter. There are several possible answers. And, depending on our answer, there are several further implications. Thus there are several moral questions which I cannot even mention here. I shall end with another remark about our concern for our own future.

That concern is of several kinds. We may want to survive so that our hopes and ambitions will be achieved. We may also care about our future in the same kind of way in which we care about the futures of certain other people, such as our relatives or friends. But most of us have, in addition, a distinctive kind of egoistic concern. If I know that my child will be in pain, I may be more concerned than I would be about my own future pain. But I cannot fearfully anticipate my child's pain.

This kind of instinctive concern may, I believe, be weakened, and be seen to have no ground, if we come to accept a reductionist view. In our thoughts about our own identity, we are prey to illusions. That is why the so-called "problem cases" seem to raise problems: why we find it hard to believe that, when we know the other facts, it is an empty or a merely verbal question whether we shall still exist. Even after we accept a reductionist view, we may continue, at some level, to think and feel as if that view were not true. Our own continued existence may still seem to be an independent fact, of a peculiarly deep and simple kind. And that belief may underlie our anticipatory concern about our own future.

There are, I suspect, several causes of that illusory belief. I have discussed one cause today: our conceptual scheme. We can mistake conceptual facts for facts about reality. And, in the case of certain concepts, those that are most loaded with emotional or moral significance, we can be led seriously astray. Of these loaded concepts, that of our own identity is, perhaps, the most misleading.

Even the use of the word "I" can lead us astray. Consider the fact that, in a few years, I shall be dead. This fact can seem depressing. But the reality is only this. After a certain time, none of the thoughts and experiences that occur will be causally dependent on this brain, or be related in certain ways to this series of experiences. That is all this fact involves. And, in that redescription, my death seems to disappear.

The author is a senior research fellow at All Souls College, Oxford. This article is a shortened extract from the second of this year's Herbert Spencer Lectures. Derek Parfit's account of personal identity is developed further in his book Reasons and Persons, *Oxford University Press, 1984.*

Next week, Henry Harris looks at experimental evidence and identity.