Reductionism and Personal Identity[1]

Derek Parfit

We can start with some science fiction. Here on Earth, I enter the Teletransporter. When I press some button, a machine destroys my body, while recording the exact states of all my cells. The information is sent by radio to Mars, where another machine makes, out of organic materials, a perfect copy of my body. The person who wakes up on Mars seems to remember living my life up to the moment when I pressed the button, and he is in every other way just like me.

Of those who have thought about such cases, some believe that it would be I who would wake up on Mars. They regard Teletransportation as merely the fastest way of travelling. Others believe that, if I chose to be Teletransported, I would be making a terrible mistake. On their view, the person who wakes up would be a mere Replica of me.

I

That is a disagreement about personal identity. To understand such disagreements, we must distinguish two kinds of sameness. Two white billiard balls may be qualitatively identical, or exactly similar. But they are not numerically identical, or one and the same ball. If I paint one of these balls red, it will cease to be qualitatively identical with itself as it was; but it will still be one and the same ball. Consider next a claim like, 'Since her accident, she is no longer the same person.' That involves both senses of identity. It means that *she,* one and the same person, is *not* now the same person. That is not a contradiction. The claim is only that this person's character has changed. This numerically identical person is now qualitatively different.

When psychologists discuss identity, they are typically concerned with the kind of person someone is, or wants to be. That is the question involved, for example, in an identity crisis. But, when philosophers discuss identity, it is numerical identity they mean. And, in our concern about our own futures, that is what we have in mind. I may believe that, after my marriage, I shall be a different person. But that does not make marriage death. However much I change, I shall still be alive if there will be someone living who will be me. Similarly, if I was Teletransported, my Replica on Mars would be qualitatively identical to me; but, on the sceptic's view, he wouldn't *be* me. *I* shall have ceased to exist. And that, we naturally assume, is what matters.

Questions about our numerical identity all take the following form. We have two ways of referring to a person, and we ask whether these are ways of referring to the same person. Thus we might ask whether Boris Nikolayevich is Yeltsin. In the most important questions of this kind, our two ways of referring to a person pick out a person at different times. Thus we might ask whether the person to whom we are speaking now is the same as the person to whom we spoke on the telephone yesterday. These are questions about identity over time.

To answer such questions, we must know the *criterion* of personal identity: the relation between a person at one time, and a person at another time, which makes these one and the same person.

Different criteria have been advanced. On one view, what makes me the same, throughout my life, is my having the same body. This criterion requires uninterrupted bodily continuity. There is no such continuity between my body on Earth and the body of my Replica on Mars; so, on this view, my Replica would not be me. Other writers appeal to psychological continuity. Thus Locke claimed that, if I was conscious of a past life in some other body, I would be the person who lived that life. On some versions of this view, my Replica would be me.

Supporters of these different views often appeal to cases where they conflict. Most of these cases are, like Teletransportation, purely imaginary. Some philosophers object that, since our concept of a person rests on a scaffolding of facts, we should not expect this concept to apply in imagined cases where we think those facts away. I agree. But I believe that, for a different reason, it is worth considering such cases. We

Excerpted from "The Unimportance of Identity," in *Identity,* edited by H. Harris (1995), pp. 13–28, with permission from Oxford University Press.

can use them to discover, not what the truth is, but what we believe. We might have found that, when we consider science fiction cases, we simply shrug our shoulders. But that is not so. Many of us find that we have certain beliefs about what kind of fact personal identity is.

These beliefs are best revealed when we think about such cases from a first-person point of view. So, when I imagine something's happening to me, you should imagine its happening to you. Suppose that I live in some future century, in which technology is far advanced, and I am about to undergo some operation. Perhaps my brain and body will be remodelled, or partially replaced. There will be a resulting person, who will wake up tomorrow. I ask, 'Will that person be me? Or am I about to die? Is this the end?' I may not know how to answer this question. But it is natural to assume that there must *be* an answer. The resulting person, it may seem, must be either me, or someone else. And the answer must be all-or-nothing. That person cannot be *partly* me. If that person is in pain tomorrow, this pain cannot be partly mine. So, we may assume, either I shall feel that pain, or I shan't.

If this is how we think about such cases, we assume that our identity must be *determinate*. We assume that, in every imaginable case, questions about our identity must have answers, which must be either, and quite simply, Yes or No.

Let us now ask: 'Can this be true?' There is one view on which it might be. On this view, there are immaterial substances: souls, or Cartesian Egos. These entities have the special properties once ascribed to atoms: they are indivisible, and their continued existence is, in its nature, all or nothing. And such an Ego is what each of us really is.

Unlike several writers, I believe that such a view might have been true. But we have no good evidence for thinking that it is, and some evidence for thinking that it isn't; so I shall assume here that no such view is true.

If we do not believe that there are Cartesian Egos, or other such entities, we should accept the kind of view which I have elsewhere called *Reductionist*. On this view

(1) A person's existence just consists in the existence of a body, and the occurrence of a series of thoughts, experiences, and other mental and physical events.

Some Reductionists claim

(2) Persons just *are* bodies.

This view may seem not to be Reductionist, since it does not reduce persons to something else. But that is only because it is hyper-Reductionist: it reduces persons to bodies in so strong a way that it doesn't even distinguish between them. We can call it *Identifying* Reductionism.

Such a view seems to me too simple. I believe that we should combine (1) with

(3) A person is an entity that has a body, and has thoughts and other experiences.

On this view, though a person is distinct from that person's body, and from any series of thoughts and experiences, the person's existence just *consists* in them. So we can call this view *Constitutive* Reductionism.

It may help to have other examples of this kind of view. If we melt down a bronze statue, we destroy this statue, but we do not destroy this lump of bronze. So, though the statue just consists in the lump of bronze, these cannot be one and the same thing. Similarly, the existence of a nation just consists in the existence of a group of people, on some territory, living together in certain ways. But the nation is not the same as that group of people, or that territory.

Consider next *Eliminative* Reductionism. Such a view is sometimes a response to arguments against the Identifying view. Suppose we start by claiming that a nation just is a group of people on some territory. We are then persuaded that this cannot be so: that the concept of a nation is the concept of an entity that is distinct from its people and its territory. We may conclude that, in that case, there are really no such things as nations. There are only groups of people, living together in certain ways.

In the case of persons, some Buddhist texts take an Eliminative view. According to these texts

(4) There really aren't such things as persons: there are only brains and bodies, and thoughts and other experiences.

For example:

> Buddha has spoken thus: 'O brethren, actions do exist, and also their consequences, but the person that acts does not. . . . There exists no Individual, it is only a conventional name given to a set of elements.'

Or:

> *The mental and the material are really here,*

But here there is no person to be found.
For it is void and merely fashioned like a doll,
Just suffering piled up like grass and sticks.

Eliminative Reductionism is sometimes justified. Thus we are right to claim that there were really no witches, only persecuted women. But Reductionism about some kind of entity is not often well expressed with the claim that there are no such entities. We should admit that there are nations, and that we, who are persons, exist.

Rather than claiming that there are no entities of some kind, Reductionists should distinguish kinds of entity, or ways of existing. When the existence of an X just consists in the existence of a Y, or Ys, though the X is *distinct* from the Y or Ys, it is not an *independent* or *separately existing* entity. Statues do not exist separately from the matter of which they are made. Nor do nations exist separately from their citizens and their territory. Similarly, I believe,

(5) Though persons are distinct from their bodies, and from any series of mental events, they are not independent or separately existing entities.

Cartesian Egos, if they existed, would not only be distinct from human bodies, but would also be independent entities. Such Egos are claimed to be like physical objects, except that they are wholly mental. If there were such entities, it would make sense to suppose that they might cease to be causally related to some body, yet continue to exist. But, on a Reductionist view, persons are not in that sense independent from their bodies. (That is not to claim that our thoughts and other experiences are merely changes in the states of our brains. Reductionists, while not believing in purely mental substances, may be dualists.)

We can now return to personal identity over time, or what constitutes the continued existence of the same person. One question here is this. What explains the unity of a person's mental life? What makes thoughts and experiences, had at different times, the thoughts and experiences of a single person? According to some Non-Reductionists, this question cannot be answered in other terms. We must simply claim that these different thoughts and experiences are all had by the same person. This fact does not consist in any other facts, but is a bare or ultimate truth.

If each of us was a Cartesian Ego, that might be so. Since such an Ego would be an independent substance, it could be an irreducible fact that different experiences are all changes in the states of the same persisting Ego. But that could not be true of persons, I believe, if, while distinct from their bodies, they are not separately existing entities. A person, so conceived, is not the kind of entity about which there could be such irreducible truths. When experiences at different times are all had by the same person, this fact must consist in certain other facts.

If we do not believe in Cartesian Egos, we should claim

(6) Personal identity over time just consists in physical and/or psychological continuity.

That claim could be filled out in different ways. On one version of this view, what makes different experiences the experiences of a single person is their being either changes in the states of, or at least directly causally related to, the same embodied brain. That must be the view of those who believe that persons just are bodies. And we might hold that view even if, as I think we should, we distinguish persons from their bodies. But we might appeal, either in addition or instead, to various psychological relations between different mental states and events, such as the relations involved in memory, or in the persistence of intentions, desires, and other psychological features. That is what I mean by psychological continuity.

On Constitutive Reductionism, the fact of personal identity is distinct from these facts about physical and psychological continuity. But, since it just consists in them, it is not an independent or separately obtaining fact. It is not a further difference in what happens.

To illustrate that distinction, consider a simpler case. Suppose that I already know that several trees are growing together on some hill. I then learn that, because that is true, there is a copse on this hill. That would not be new factual information. I would have merely learnt that such a group of trees can be called a 'copse.' My only new information is about our language. That those trees can be called a copse is not, except trivially, a fact about the trees.

Something similar is true in the more complicated case of nations. In order to know the facts about the history of a nation, it is enough to know what large numbers of people did and said. Facts about nations cannot be barely true:

they must consist in facts about people. And, once we know these other facts, any remaining questions about nations are not further questions about what really happened.

I believe that, in the same way, facts about people cannot be barely true. Their truth must consist in the truth of facts about bodies, and about various interrelated mental and physical events. If we knew these other facts, we would have all the empirical input that we need. If we understood the concept of a person, and had no false beliefs about what persons are, we would then know, or would be able to work out, the truth of any further claims about the existence or identity of persons. That is because such claims would not tell us more about reality.

That is the barest sketch of a Reductionist view. These remarks may become clearer if we return to the so-called 'problem cases' of personal identity. In such a case, we imagine knowing that, between me now and some person in the future, there will be certain kinds or degrees of physical and/or psychological continuity or connectedness. But, though we know these facts, we cannot answer the question whether that future person would be me.

Since we may disagree on which the problem cases are, we need more than one example. Consider first the range of cases that I have elsewhere called the *Physical Spectrum.* In each of these cases, some proportion of my body would be replaced, in a single operation, with exact duplicates of the existing cells. In the case at the near end of this range, no cells would be replaced. In the case at the far end, my whole body would be destroyed and replicated. That is the case with which I began: Teletransportation.

Suppose we believe that in that case, where my whole body would be replaced, the resulting person would not be me, but a mere Replica. If no cells were replaced, the resulting person would be me. But what of the cases in between, where the percentage of the cells replaced would be, say, 30, or 50, or 70 per cent? Would the resulting person here be me? When we consider some of these cases, we will not know whether to answer Yes or No.

Suppose next that we believe that, even in Teletransportation, my Replica would be me. We should then consider a different version of that case, in which the Scanner would get its information without destroying my body, and my Replica would be made while I was still alive. In this version of the case, we may agree that my Replica would not be me. That may shake our view that, in the original version of case, he *would* be me.

If we still keep that view, we should turn to what I have called the *Combined Spectrum.* In this second range of cases, there would be all the different degrees of both physical and psychological connectedness. The new cells would not be exactly similar. The greater the proportion of my body that would be replaced, the less like me would the resulting person be. In the case at the far end of this range, my whole body would be destroyed, and they would make a Replica of some quite different person, such as Greta Garbo. Garbo's Replica would clearly *not* be me. In the case at the near end, with no replacement, the resulting person would be me. On any view, there must be cases in between where we could not answer our question.

For simplicity, I shall consider only the Physical Spectrum, and I shall assume that, in some of the cases in this range, we cannot answer the question whether the resulting person would be me. My remarks could be transferred, with some adjustment, to the Combined Spectrum.

As I have said, it is natural to assume that, even if *we* cannot answer this question, there must always *be* an answer, which must be either Yes or No. It is natural to believe that, if the resulting person will be in pain, either I shall feel that pain, or I shan't. But this range of cases challenges that belief. In the case at the near end, the resulting person would be me. In the case at the far end, he would be someone else. How could it be true that, in all the cases in between, he must be either me, or someone else? For that to be true, there must be, somewhere in this range, a sharp borderline. There must be some critical set of cells such that, if only those cells were replaced, it would be me who would wake up, but that in the very next case, with only just a few more cells replaced, it would be, not me, but a new person. That is hard to believe.

Here is another fact, which makes it even harder to believe. Even if there were such a borderline, no one could ever discover where it is. I might say, 'Try replacing half of my brain and body, and I shall tell you what happens.' But we know in advance that, in every case, since the resulting person would be exactly like me, he would be inclined to believe that he was me. And this could not show that he *was* me, since any mere Replica of me would think that too.

Even if such cases actually occurred, we would learn nothing more about them. So it does not matter that these cases are imaginary.

We should try to decide now whether, in this range of cases, personal identity could be determinate. Could it be true that, in every case, the resulting person either would or would not be me?

If we do not believe that there are Cartesian Egos, or other such entities, we seem forced to answer No. It is not true that our identity must be determinate. We can always ask, 'Would that future person be me?' But, in some of these cases,

(7) This question would have no answer. It would be neither true nor false that this person would be me.

And

(8) This question would be *empty*. Even without an answer, we could know the full truth about what happened.

If our questions were about such entities as nations or machines, most of us would accept such claims. But, when applied to ourselves, they can be hard to believe. How could it be neither true nor false that I shall still exist tomorrow? And, without an answer to our question, how could I know the full truth about my future?

Reductionism gives the explanation. We naturally assume that, in these cases, there are different possibilities. The resulting person, we assume, might be me, or he might be someone else, who is merely like me. If the resulting person will be in pain, either I shall feel that pain, or I shan't. If these really were different possibilities, it would be compelling that one of them must be the possibility that would in fact obtain. How could reality fail to choose between them? But, on a Reductionist view,

(9) Our question is not about different possibilities. There is only a single possibility, or course of events. Our question is merely about different possible descriptions of this course of events.

That is how our question has no answer. We have not yet decided which description to apply. And, that is why, even without answering this question, we could know the full truth about what would happen.

Suppose that, after considering such examples, we cease to believe that our identity must be determinate. That may seem to make little difference. It may seem to be a change of view only about some imaginary cases, that will never actually occur. But that may not be so. We may be led to revise our beliefs about the nature of personal identity; and that would be a change of view about our own lives.

In nearly all actual cases, questions about personal identity have answers, so claim (7) does not apply. If we don't know these answers, there is something that we don't know. But claim (8) still applies. Even without answering these questions, we could know the full truth about what happens. We would know that truth if we knew the facts about both physical and psychological continuity. If, implausibly, we still didn't know the answer to a question about identity, our ignorance would only be about our language. And that is because claim (9) still applies. When we know the other facts, there are never different possibilities at the level of what happens. In all cases, the only remaining possibilities are at the linguistic level. Perhaps it would be correct to say that some future person would be me. Perhaps it would be correct to say that he would not be me. Or perhaps neither would be correct. I conclude that in *all* cases, if we know the other facts, we should regard questions about our identity as merely questions about language.

That conclusion can be misunderstood. First, when we ask such questions, that is usually because we *don't* know the other facts. Thus, when we ask if we are about to die, that is seldom a conceptual question. We ask that question because we don't know what will happen to our bodies, and whether, in particular, our brains will continue to support consciousness. Our question becomes conceptual only when we already know about such other facts.

Note next that, in certain cases, the relevant facts go beyond the details of the case we are considering. Whether some concept applies may depend on facts about other cases, or on a choice between scientific theories. Suppose we see something strange happening to an unknown animal. We might ask whether this process preserves the animal's identity, or whether the result is a new animal (because what we are seeing is some kind of reproduction). Even if we knew the details of this process, that question would not be merely conceptual. The answer would depend on whether this process is part of the natural development of this kind of animal. And that may be something we have yet to discover.

If we identify persons with human beings, whom we regard as a natural kind, the same

would be true in some imaginable cases involving persons. But these are not the kind of case that I have been discussing. My cases all involve artificial intervention. No facts about natural development could be relevant here. Thus, in my Physical Spectrum, if we knew which of my cells would be replaced by duplicates, all of the relevant empirical facts would be in. In such cases any remaining questions would be conceptual.

Since that is so, it would be clearer to ask these questions in a different way. Consider the case in which I replace some of the components of my audio system, but keep the others. I ask, 'Do I still have one and the same system?' That may seem a factual question. But, since I already know what happened, that is not really so. It would be clearer to ask, 'Given that I have replaced those components, would it be correct to call this the same system?'

The same applies to personal identity. Suppose that I know the facts about what will happen to my body, and about any psychological connections that there will be between me now and some person tomorrow. I may ask, 'Will that person be me?' But that is a misleading way to put my question. It suggests that I don't know what's going to happen. When I know these other facts, I should ask, 'Would it be correct to call that person me?' That would remind me that, if there's anything that I don't know, that is merely a fact about our language.

I believe that we can go further. Such questions are, in the belittling sense, merely verbal. Some conceptual questions are well worth discussing. But questions about personal identity, in my kind of case, are like questions that we would all think trivial. It is quite uninteresting whether, with half its components replaced, I still have the same audio system. In the same way, we should regard it as quite uninteresting whether, if half of my body were simultaneously replaced, I would still exist. As questions about reality, these are entirely empty. Nor, as conceptual questions, do they need answers.

We might need, for legal purposes, to *give* such questions answers. Thus we might decide that an audio system should be called the same if its new components cost less than half its original price. And we might decide to say that I would continue to exist as long as less than half my body were replaced. But these are not answers to conceptual questions; they are mere decisions.

(Similar remarks apply if we are Identifying Reductionists, who believe that persons just are bodies. There are cases where it is a merely verbal question whether we still have one and the same human body. That is clearly true in the cases in the middle of the Physical Spectrum.)

It may help to contrast these questions with one that is not merely verbal. Suppose we are studying some creature which is very unlike ourselves, such as an insect, or some extraterrestrial being. We know all the facts about this creature's behaviour, and its neurophysiology. The creature wriggles vigorously, in what seems to be a response to some injury. We ask, 'Is it conscious, and in great pain? Or is it merely like an insentient machine?' Some Behaviourist might say, 'That is a merely verbal question. These aren't different possibilities, either of which might be true. They are merely different descriptions of the very same state of affairs.' That I find incredible. These descriptions give us, I believe, two quite different possibilities. It could not be an empty or a merely verbal question whether some creature was unconscious or in great pain.

It is natural to think the same about our own identity. If I know that some proportion of my cells will be replaced, how can it be a merely verbal question whether I am about to die, or shall wake up again tomorrow? It is because that is hard to believe that Reductionism is worth discussing. If we become Reductionists, that may change some of our deepest assumptions about ourselves.

These assumptions, as I have said, cover actual cases, and our own lives. But they are best revealed when we consider the imaginary problem cases. It is worth explaining further why that is so.

In ordinary cases, questions about our identity have answers. In such cases, there is a fact about personal identity, and Reductionism is one view about what kind of fact this is. On this view, personal identity just consists in physical and/or psychological continuity. We may find it hard to decide whether we accept this view, since it may be far from clear when one fact just consists in another. We may even doubt whether Reductionists and their critics really disagree.

In the problem cases, things are different. When we cannot answer questions about personal identity, it is easier to decide whether we accept a Reductionist view. We should ask: Do we find such cases puzzling? Or do we accept the Reductionist claim that, even without answering these questions, if we knew the facts

about the continuities, we would know what happened?

Most of us do find such cases puzzling. We believe that, even if we knew those other facts, if we could not answer questions about our identity, there would be something that we didn't know. That suggests that, on our view, personal identity does *not* just consist in one or both of the continuities, but is a separately obtaining fact, or a further difference in what happens. The Reductionist account must then leave something out. So there is a real disagreement, and one that applies to all cases.

Many of us do not merely find such cases puzzling. We are inclined to believe that, in all such cases, questions about our identity must have answers, which must be either Yes or No. For that to be true, personal identity must be a separately obtaining fact of a peculiarly simple kind. It must involve some special entity, such as a Cartesian Ego, whose existence must be all-or-nothing.

When I say that we have these assumptions, I am *not* claiming that we believe in Cartesian Egos. Some of us do. But many of us, I suspect, have inconsistent beliefs. If we are asked whether we believe that there are Cartesian Egos, we may answer No. And we may accept that, as Reductionists claim, the existence of a person just involves the existence of a body, and the occurrence of a series of interrelated mental and physical events. But, as our reactions to the problem cases show, we don't fully accept that view. Or, if we do, we also seem to hold a different view.

Such a conflict of beliefs is quite common. At a reflective or intellectual level, we may be convinced that some view is true; but at another level, one that engages more directly with our emotions, we may continue to think and feel as if some different view were true. One example of this kind would be a hope, or fear, that we know to be groundless. Many of us, I suspect, have such inconsistent beliefs about the metaphysical questions that concern us most, such as free will, time's passage, consciousness, and the self. . . .

NOTE

1. Some of this essay draws from Part Three of my *Reasons and Persons* (Oxford University Press, 1984). The new material will be more fully developed in my contribution to Dancy, *Derek Parfit and His Critics: Vol. I. Persons* (Blackwell's, forthcoming).