2

The Unimportance of Identity

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 ident-

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).

ical, 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 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 physi-

cal 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.

II

I turn now from the nature of personal identity to its importance. Personal identity is widely thought to have great rational and moral significance. Thus it is the fact of identity which is thought to give us our reason for concern about our own future. And several moral principles, such as those of desert or distributive justice, presuppose claims about identity. The separateness of persons, or the non-identity of different people, has been called 'the basic fact for morals'.

I can comment here on only one of these questions: what matters in our survival. I mean by that, not what makes our survival good, but what makes our survival matter, whether it will be good or bad. What is it, in our survival, that gives us a reason for special anticipatory or prudential concern?

We can explain that question with an extreme imaginary case. Suppose that, while I care about my whole future, I am especially concerned about what will happen to me on future Tuesdays. Rather than suffer mild pain on a future Tuesday, I would choose severe pain on any other future day. That pattern of concern would be irrational. The fact that a pain will be on a Tuesday is no reason to care about it more. What about the fact that a pain will be *mine*? Is *this* a reason to care about it more?

Many people would answer Yes. On their view, what gives us a reason to care about our future is, precisely, that it will be our future. Personal identity is what matters in survival.

I reject this view. Most of what matters, I believe, are two other relations: the psychological continuity and connectedness that, in ordinary cases, hold between the different parts of a person's life. These relations only roughly coincide with personal identity, since, unlike identity, they are in part matters of degree. Nor, I believe, do they matter as much as identity is thought to do.

There are different ways to challenge the importance of identity.

One argument can be summarized like this:

(1) Personal identity just consists in certain other facts.
(2) If one fact just consists in certain others, it can only be these other facts which have rational or moral importance. We should ask whether, in themselves, these other facts matter.

Therefore

(3) Personal identity cannot be rationally or morally important. What matters can only be one or more of the other facts in which personal identity consists.

Mark Johnston rejects this argument.[1] He calls it an *Argument from Below*, since it claims that, if one fact justs consists in certain others, it can only be these other lower level facts which matter. Johnston replies with what he calls an *Argument from Above*. On his view, even if the lower-level facts do not in themselves matter, the higher-level fact may matter. If it does, the lower-level facts will have a derived significance. They will matter, not in themselves, but because they constitute the higher-level fact.

To illustrate this disagreement, we can start with a different case. Suppose we ask what we want to happen if, through

[1] In his 'Human Concerns Without Superlative Selves', in Dancy (forthcoming).

brain damage, we become irreversibly unconscious. If we were in this state, we would still be alive. But this fact should be understood in a Reductionist way. It may not be the same as the fact that our hearts would still be beating, and our other organs would still be functioning. But it would not be an independent or separately obtaining fact. Our being still alive, though irreversibly unconscious, would just consist in these other facts.

On my Argument from Below, we should ask whether those other facts in themselves matter. If we were irreversibly unconscious, would it be either good for us, or good for others, that our hearts and other organs would still be functioning? If we answer No, we should conclude that it would not matter that we were still alive.

If Johnston were right, we could reject this argument. And we could appeal to an Argument from Above. We might say:

> It may not be in itself good that our hearts and other organs would still be functioning. But it is good to be alive. Since that is so, it is rational to hope that, even if we could never regain consciousness, our hearts would go on beating for as long as possible. That would be good because it would constitute our staying alive.

I believe that, of these arguments, mine is more plausible.

Consider next the moral question that such cases raise. Some people ask, in their living wills, that if brain damage makes them irreversibly unconscious, their hearts should be stopped. I believe that we should do what these people ask. But many take a different view. They could appeal to an Argument from Above. They might say:

> Even if such people can never regain consciousness, while their hearts are still beating, they can be truly called alive. Since that is so, stopping their hearts would be an act of killing. And, except in self-defence, it is always wrong to kill.

On this view, we should leave these people's hearts to go on beating, for months or even years.

As an answer to the moral question, this seems to me misguided. (It is a separate question what the law should be.)

But, for many people, the word 'kill' has such force that it seems significant whether it applies.

Turn now to a different subject. Suppose that, after trying to decide when people have free will, we become convinced by either of two compatibilist views. On one view, we call choices 'unfree' if they are caused in certain ways, and we call them 'free' if they are caused in certain other ways. On the other view, we call choices 'unfree' if we know how they were caused, and we call them 'free' if we have not yet discovered this.

Suppose next that, when we consider these two grounds for drawing this distinction, we believe that neither, in itself, has the kind of significance that could support making or denying claims about guilt, or desert. There seems to us no such significance in the difference between these kinds of causal determination; and we believe that it cannot matter whether a decision's causes have already been discovered. (Note that, in comparing the Arguments from Above and Below, we need not actually accept these claims. We are asking whether, *if* we accepted the relevant premisses, we ought to be persuaded by these arguments.)

On my Argument from Below, if the fact that a choice is free just consists in one of those other facts, and we believe that those other facts cannot in themselves be morally important, we should conclude that it cannot be important whether some person's choice was free. Either choices that are unfree can deserve to be punished, or choices that are free cannot. On a Johnstonian Argument from Above, even if those other facts are not in themselves important—even if, in themselves, they are trivial—they can have a derived importance if and because they constitute the fact that some person's choice was free. As before, the Argument from Below seems to me more plausible.

We can now consider the underlying question on which this disagreement turns.

As I have claimed, if one fact just consists in certain others, the first fact is not an independent or separately obtaining

fact. And, in the cases with which we are concerned, it is also, in relation to these other facts, merely a conceptual fact. Thus, if someone is irreversibly unconscious, but his heart is still beating, it is a conceptual fact that this person is still alive. When I call this fact conceptual, I don't mean that it is a fact about our concepts. That this person is alive is a fact about this person. But, if we have already claimed that this person's heart is still beating, when we claim that he is still alive, we do not give further information about reality. We only give further information about our use of the words 'person' and 'alive'.

When we turn to ask what matters, the central question is this. 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 fact does matter, because it makes it correct to say that this person is still alive? 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.

This, I believe, is irrational. On my view, what matters are the facts about the world, given which some concept applies. If the facts about the world have no rational or moral significance, and the fact that the concept applies is not a further difference in what happens, this conceptual fact cannot be significant.

Johnston brings a second charge against my argument. If physicalism were true, he claims, all facts would just consist in facts about fundamental particles. Considered in themselves, these facts about particles would have no rational or moral importance. If we apply an Argument from Below, we must conclude that nothing has any importance. He remarks: 'this is not a proof of Nihilism. It is a reductio ad absurdum.'

Given what I have suggested here, this charge can, I think, be answered. There may perhaps be a sense in which, if physicalism were true, all facts would just consist in facts

about fundamental particles. But that is not the kind of reduction which I had in mind. When I claim that personal identity just consists in certain other facts, I have in mind a closer and partly conceptual relation. Claims about personal identity may not mean the same as claims about physical and/or psychological continuity. But, if we knew the facts about these continuities, and understood the concept of a person, we would thereby know, or would be able to work out, the facts about persons. Hence my claim that, if we know the other facts, questions about personal identity should be taken to be questions, not about reality, but only about our language. These claims do not apply to facts about fundamental particles. It is not true for example that, if we knew how the particles moved in some person's body, and understood our concepts, we would thereby know, or be able to work out, all of the relevant facts about this person. To understand the world around us, we need more than physics and a knowledge of our own language.

My argument does not claim that, whenever there are facts at different levels, it is always the lowest-level facts which matter. That is clearly false. We are discussing cases where, relative to the facts at some lower level, the higher-level fact is, in the sense that I have sketched, merely conceptual. My claim is that such conceptual facts cannot be rationally or morally important. What matters is reality, not how it is described. So this view might be called *realism about importance.*

If we are Reductionists about persons, and Realists about importance, we should conclude that personal identity is not what matters. Can we accept that conclusion?

Most of us believe that we should care about our future because it will be *our* future. I believe that what matters is not identity but certain other relations. To help us to decide between these views, we should consider cases where identity and those relations do not coincide.

Which these cases are depends on which criterion of identity we accept. I shall start with the simplest form of the

Physical Criterion, according to which a person continues to exist if and only if that person's body continues to exist. That must be the view of those who believe that persons just are bodies. And it is the view of several of the people who identify persons with human beings. Let's call this the *Bodily Criterion.*

Suppose that, because of damage to my spine, I have become partly paralysed. I have a brother, who is dying of a brain disease. With the aid of new techniques, when my brother's brain ceases to function, my head could be grafted onto the rest of my brother's body. Since we are identical twins, my brain would then control a body that is just like mine, except that it would not be paralysed.

Should I accept this operation? Of those who assume that identity is what matters, three groups would answer No. Some accept the Bodily Criterion. These people believe that, if this operation were performed, I would die. The person with my head tomorrow would be my brother, who would mistakenly think that he was me. Other people are uncertain what would happen. They believe that it would be risky to accept this operation, since the resulting person might not be me. Others give a different reason why I should reject this operation: that it would be indeterminate whether that person would be me. On all these views, it matters who that person would be.

On my view, that question is unimportant. 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 like me. These facts would also have their normal cause, the continued existence of my brain. And this person's body would be just like mine. For all these reasons, his life would be just like the life that I would have lived, if my paralysis had been cured. I believe that, given these facts, I should accept this operation. It is irrelevant whether this person would be me.

That may seem all important. After all, if he would not be me, I shall have ceased to exist. But, if that person would not

be me, this fact would just consist in another fact. It would just consist in the fact that my body will have been replaced below the neck. When considered on its own, is that second fact important? Can it matter in itself that the blood that will keep my brain alive will circulate, not through my own heart and lungs, but through my brother's heart and lungs? Can it matter in itself that my brain will control, not the rest of my body, but the rest of another body that is exactly similar?

If we believe that these facts would amount to my non-existence, it may be hard to focus on the question whether, in themselves, these facts matter. To make that easier, we should imagine that we accept a different view. Suppose we are convinced that the person with my head tomorrow *would* be me. Would we then believe that it would matter greatly that my head would have been grafted onto this other 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. As this shows, if it would matter greatly that what will be replaced is not just a few such organs, but my whole body below the neck, that could only be because, if that happened, the resulting person would *not* be me.

According to my argument, we should now conclude that neither of these facts could matter greatly. Since it would not be in itself important that my head would be grafted onto this body, and that would be all there was to the fact that the resulting person would not be me, it would not be in itself important that this person would not be me. Perhaps it would not be irrational to regret these facts a little. But, I believe, they would be heavily outweighed by the fact that, unlike me, the resulting person would not be paralysed.

When it is applied to our own existence, my argument is hard to accept. But, as before, the fundamental question is the relative importance of language and reality.

On my view, what matters is what is going to happen. If I knew that my head could be grafted onto the rest of a body that is just like mine, and that the resulting person would be

just like me, I would know enough to decide whether to accept this operation. I need not ask whether the resulting person could be correctly called me. That is not a further difference in what is going to happen.

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 doctor whether, while I receive some treatment, I shall be in pain. That is a factual question. I am asking what will happen. 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, while I am crossing the Channel, I shall be feeling 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 am crossing the Channel, shall I be in pain?' But that would be misleading, since it would suggest that I am asking what will happen.

In the medical 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 medical treatment, I shall be in pain. And it matters whether, while crossing the Channel, I shall be sea-sick. But it does not matter whether, in feeling sea-sick, I can be said to be in pain.

Return now to our main example. Suppose I know that my head will be successfully grafted onto my brother's headless body. I ask whether the resulting person will be me. Is this

like the medical case, or the case of crossing the Channel? Am I asking what will happen, or whether what I know will happen could be described in a certain way?

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

It may now be objected: 'By choosing this example, you are cheating. Of course you should accept this operation. But that is because the resulting person *would* be you. We should reject the Bodily Criterion. So this case cannot show that identity is not what matters.'

Since there are people who accept this criterion, I am not cheating. It is worth trying to show these people that identity is not what matters. But I accept part of this objection. I agree that we should reject the Bodily Criterion.

Of those who appeal to this criterion, some believe that persons just are bodies. But, if we hold this kind of view, it would be better to identify a person with that person's brain, or nervous system. Consider next those who believe that persons are animals of a certain kind, viz. human beings. We could take this view, but reject the Bodily Criterion. We could claim that animals continue to exist if there continue to exist, and to function, the most important parts of their bodies. And we could claim that, at least in the case of human beings, the brain is so important that its survival counts as the survival of this human being. On both these views, in my imagined case, the person with my head tomorrow would be me. And that is what, on reflection, most of us would believe.

My own view is similar. 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.

If we agree that this person would be me, I would still argue that this fact is not what matters. What is important is

not identity, but one or more of the other facts in which identity consists. But I concede that, when identity coincides with these other facts, it is harder to decide whether we accept that argument's conclusion. So, if we reject the Bodily Criterion, we must consider other cases.

Suppose that we accept the Brain-Based version of the Psychological Criterion. On this view, if there will be one future person who is psychologically continuous with me, because he will have enough of my brain, that person will be me. But psychological continuity without its normal cause, the continued existence of enough of my brain, does not suffice for identity. My Replica would not be me.

Remember next that an object can continue to exist even if all its components are gradually replaced. Suppose that, every time some wooden ship comes into port, a few of its planks are replaced. Before long, the same ship may be entirely composed of different planks.

Assume, once again, that I need surgery. All of my brain cells have a defect which, in time, would be fatal. Surgeons could replace all these cells, inserting new cells that are exact replicas, except that they have no defect.

The surgeons could proceed in either of two ways. In *Case One*, there would be a hundred operations. In each operation, the surgeons would remove a hundredth part of my brain, and insert replicas of those parts. In *Case Two*, the surgeons would first remove all the existing parts of my brain and then insert all of their replicas.

There is a real difference here. In Case One, my brain would continue to exist, like a ship with all of its planks gradually replaced. In Case Two, my brain would cease to exist, and my body would be given a new brain.

This difference, though, is much smaller than that between ordinary survival and teletransportation. In both cases, there will later be a person whose brain will be just like my present brain, but without the defects, and who will therefore be psychologically continuous with me. And, in *both* cases, this person's brain will be made of the very same new cells, each

of which is a replica of one of my existing cells. The difference between the cases is merely the way in which these new cells are inserted. In Case One, the surgeons alternate between removing and inserting. In Case Two, they do all the removing before all the inserting.

On the Brain-Based Criterion, this is the difference between life and death. In Case One, the resulting person would be me. In Case Two he would *not* be me, so I would cease to exist.

Can this difference matter? Reapply the Argument from Below. This difference consists in the fact that, rather than alternating between removals and insertions, the surgeon does all the removing before all the inserting. Considered on its own, can this matter? I believe not. We would not think it mattered if it did not constitute the fact that the resulting person would not be me. But if this fact does not in itself matter, and that is all there is to the fact that in Case Two I would cease to exist, I should conclude that my ceasing to exist does not matter.

Suppose next that you regard these as problem cases, ones where you do not know what would happen to me. Return to the simpler Physical Spectrum. In each of the cases in this range, some proportion of my cells will be replaced with exact duplicates. With some proportions—20 per cent, say, or 50, or 70—most of us would be uncertain whether the resulting person would be me. (As before, if we do not believe that here, my remarks could be transferred, with adjustments, to the Combined Spectrum.)

On my view, in all of the cases in this range, it is a merely conceptual question whether the resulting person would be me. Even without answering this question, I can know just what is going to happen. If there is anything that I don't know, that is merely a fact about how we could describe what is going to happen. And that conceptual question is not even, I believe, interesting. It is merely verbal, like the question whether, if I replaced some of its parts, I would still have the same audio system.

When we imagine these cases from a first-person point of view, it may still be hard to believe that this is merely a verbal question. If I don't know whether, tomorrow, I shall still exist, it may be hard to believe that I know what is going to happen. But what is it that I don't know? If there are different possibilities, at the level of what happens, what is the difference between them? In what would that difference consist? If I had a soul, or Cartesian Ego, there might be different possibilities. Perhaps, even if *n* per cent of my cells were replaced, my soul would keep its intimate relation with my brain. Or perhaps another soul would take over. But, we have assumed, there are no such entities. What else could the difference be? When the resulting person wakes up tomorrow, what could make it either 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? 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.

We may believe that this difference needs no explanation. It may seem enough to say: Perhaps that person will be me, and perhaps he won't. 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 still 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, between now and tomorrow, my brain and body might either remain unharmed, or be blown to pieces. But, in our imagined case, there is no such other difference. I already know that there will be someone whose brain and body will consist partly of these cells, and partly of new cells,

and that this person will be psychologically like me. There aren't, at the level of what happens, different possible outcomes. There is no further essence of me, or property of meness, 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 someone else. 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 still 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?

III

I am now assuming that we accept the Brain-Based Psychological Criterion. We believe that, if there will be one future person who will have enough of my brain to be psychologically continuous with me, that person would be me. On this view, there is another way to argue that identity is not what matters.

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 psychological change, even when, through a stroke or injury, they have lost the use of half their brain.

Let us next suppose that the two halves of my brain could each fully support ordinary psychological functioning. That may in fact be true of certain people. If it is not, we can suppose that, through some technological advance, it has been made true of me. Since our aim is to test our beliefs about what matters, there is no harm in making such assumptions.

We can now compare two more 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, it would be unreasonable to deny that I would survive if that remaining half were transplanted. So, in this *Single Case*, the resulting person would be me.

Consider next the *Double Case*, or *My Division*. Both halves of my brain would be successfully transplanted, into different bodies that are just like mine. 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 cannot 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 could not 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 Single Case, the one resulting person would be me. The relation between me now and that future person is just an instance of the relation between me now and myself tomorrow. So that relation would contain what matters. In the Double 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 now and that future person is in itself the same. As in the Single Case, he has half my brain, and he is just like me. The difference is only that, in this Double 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 Single Case, where half my brain will be successfully transplanted, my prospect is survival. That prospect contains what matters. In the Double 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 are like 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 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 start to be born as triplets. Should their parents think this a disaster, because these children don't have twins? Clearly not. These children don't 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 them will suffer damage as an only child. These people should revise their view. What matters isn't having a twin: it is having at least one sibling.

In the same way, we should revise our view about identity over time. 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 has 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 could not be called my twins.[2]

IV

If, as I have argued, personal identity is not 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 have no time even to mention. I shall end with another remark about our concern for our own future.

That concern is of several kinds. We may want to survive partly so that our hopes and ambitions will be achieved. We may also care about our future in the kind of way in which we care about the well-being 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 care about his pain more than I would about my own future pain. But I cannot fearfully anticipate my child's pain. And if I knew that my Replica would take up my life where I leave off, I would not look forward to that life.

[2] In many contexts, we need to distinguish two senses of 'what matters in survival'. What matters in the *prudential* sense is what gives us reason for special concern about our future. What matters in the *desirability* sense is what makes our survival good. But, in the examples I have been discussing, these two coincide. On my view, even if we won't survive, we could have what matters *in* survival. If there will be at least one living person who will both be psychological continuous with me, and have enough of my brain, my relation to that person contains what matters in the prudential sense. So it also preserves what matters in the desirability sense. It is irrelevant whether that person will be me.

This kind of 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 prone 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 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 here: our conceptual scheme. Though we need concepts to think about reality, we sometimes confuse the two. We 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 directly causally related to this brain, or be connected in certain ways to these present experiences. That is all this fact involves. And, in that redescription, my death seems to disappear.