

IS PERSONAL IDENTITY WHAT MATTERS?

by Derek Parfit

31 December 2007

In my book *Reasons and Persons*, I defended one view about the metaphysics of persons, and also claimed that personal identity is not what matters. In this paper I shall give some further arguments for this second claim, and also try to respond to some forceful objections by Mark Johnston. I shall not, however, try to decide what *does* matter, since that is a much larger and more difficult question.¹

PART ONE

1 The Metaphysics of Persons

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.

That is a disagreement about personal identity. To understand such disagreements, we must distinguish two kinds of sameness. Two 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 a different colour, 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. For example, we might ask whether Bonaparte was the same as Napoleon. In the most important questions of this kind, our two ways of referring to a person pick out a person at different times. Thus I might ask whether the white-haired man who has just entered the room is the same as the boy with whom I used to play marbles. 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 can't be *partly* me. If that person will be in pain tomorrow, this pain can't 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.

We can 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, there are actions , and also their consequences, but there is no person who acts. . . . There exists no Individual, it is only a conventional name given to a set of elements.'²

Or:

The mental and the material are really here,
 But here there is no person to be found.
 For it is void and fashioned like a doll,
 Just suffering piled up like grass and sticks. ³

Eliminative Reductionism is sometimes justified. Thus we are right to claim that there weren't really any 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.

Entities are independent, or separately existing, not absolutely, but in their relation to other entities. In relation to the plinth on which it stands, a statue is a separately existing entity. Cartesian Egos, if they existed, would be not only distinct from human bodies, but also, in relation to these bodies, 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 brain, yet continue to exist. But a statue could not exist separately from the matter of which it is made. Nor could a nation exist separately from the people in whose existence, and activities, the existence of the nation consists. In the same way, on Constitutive Reductionism, persons are distinct from their bodies, and from the mental states and events, in whose existence and occurrence their existence consists. But persons could not exist separately from these.

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. These relations together constitute what I call psychological *connectedness*, which is a matter of degree. Psychological *continuity* consists of overlapping chains of such connections.

According to 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.

These remarks could be misunderstood. First, the claim that this copse exists is not a claim *about* our language. It is a claim about reality. My point is only that, if we already know that the trees exist, learning that the copse exists isn't learning *more* about reality. Our only new information is about our language. Second, since our language is a part of reality, we do, in a way, learn more about reality. But we are not learning more about what is on the hill. 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 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 *Teletransportation*, in which 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%? Would the resulting person here be me? When we consider some of these cases, we won't 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 *Branch-Line 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 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 in which 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 can't 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* can't 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 doesn't 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.⁴ 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 ourselves, and 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.

As before, 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 know 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 brain and 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 of my brain and body were replaced. But these are not answers to conceptual questions; they are mere decisions.

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 extra-terrestrial 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 couldn't 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. We do not believe that our continued existence merely involves such continuities. It seems to us to be a further fact, of a deep and simple kind: a fact which, in every case, must be either wholly present or wholly absent. *There is no such fact.* That is my most important metaphysical claim.

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 when we think about the some of central subjects of metaphysics, such as the Self, consciousness, time's passage, and free will.

2 The Relevant Questions

After discussing the nature of personal identity, I turned to its importance. I began by asking 'What matters in survival?'⁵

I used that phrase, confusingly, to ask two different questions. One was

The prudential question: Why should we be specially concerned about our own future? What is it, in our survival, that gives us reason for such concern?

Some people believe that we have no reason for such concern. On this view, practical reasons are given by facts about our desires, and we have no reasons to have these desires. More exactly, we can have reasons to want what is a means to the fulfilment of some other desire, but we have no reason to care about anything as an end, or for its own sake. I argue against this view elsewhere.⁶

Of those who believe that we do have reasons to care about our future, most assume that personal identity provides that reason. On this view, we should be specially concerned about our future because it will be *our* future. If all we know is that someone will later be in pain, we may have some reason for concern; but, if we learn that *we* shall be in pain, we have a distinctive, further reason. And, if only one of us will be happy, we have a reason to want that person to be us. On this view, personal identity is what, in the prudential sense, matters.

This view, I argued, is mistaken. What matters in survival is not, I claimed, survival itself, but certain other relations. Though these relations usually coincide with survival, or the continued existence of the same person, that is not always so. There could be survival without what matters, and what matters without survival.

I also asked

The desirability question: What makes survival good? In wanting to survive, what do we have reason to want?

This question can be subdivided. What we have reason to want is a good future, which has the right relation to us. So we should ask:

(1) What would make our future good?

(2) If our future will be good, how should we want that future to be related to us?

In a discussion of personal identity, we need not try to answer question (1). Personal identity is not part of what makes our future good. We do need to answer (2). But (2) is not a new question; it is merely an optimistic version of the prudential question. If our future will be good, we should want its relation to us to be the one that gives us reason for special concern. As before, we must ask what that relation is.

Most people believe, as I have said, that identity provides the answer. On this view, it is enough that this future will be ours. What matters, in the desirability sense, is that *we* survive, and have a good future.

As before, I disagree. On my view, we could have what matters, in the desirability sense, even if we shall not survive. And we might have survival, and a good future, without what matters.

Though the desirability question includes the prudential question, these questions are quite different. That difference is clearest when our future will be bleak. Thus, if we knew that we would soon be paralyzed, we might not want to survive. Paralysis might destroy what, in the desirability sense, matters. But it would not destroy what prudentially matters. We would have as much reason to care about our paralyzed future.

Because I failed to distinguish these questions in my *Reasons and Persons*, I made various mistakes. Thus, when discussing what prudentially matters, I claimed that physical continuity has little value. That, even if true, is irrelevant. And, when I claimed that physical similarity might have some importance, I was again shifting, in a confused way, to the desirability question. Physical similarity is not part of what prudentially matters. If our appearance was about to change, that would not reduce our reason to care about our future.

When I discussed the imagined case which I called *My Division* and shall discuss further below, I made a similar mistake. I claimed that, if the two resulting people would have lives that were worse than mine, my relation to these people would be 'not quite the relation that matters in ordinary survival'. That again conflates our two questions. As the case of paralysis shows, even if these people would have wretched lives, that is no ground for thinking that *My Division* would destroy what prudentially matters. I might have just as much reason to be specially concerned about these people.

These mistakes can be corrected. But my argument may seem to go more seriously astray. When I discussed what prudentially matters, I sometimes appealed to imagined cases in which, if I had some operation, it would be

doubtful whether the resulting person would be me. I asked whether, in such a case, my prospect would be as good as ordinary survival. It may seem that, in asking that question, I again confused the desirability and prudential questions.

That is not so. In these imagined cases, I could either (A) have some identity-threatening operation, or (B) continue to exist in an ordinary way. I asked whether, from my present point of view, outcome (A) would be as good as (B). There is a special reason why, in these cases, that is a useful question. If I had this operation, even if the resulting person would not be me, he would be exactly like me. And this person's future would, I supposed, be very like the future that I would have in outcome (B). In such cases, the desirability and prudential questions coincide. For my prospect to contain what matters, in the desirability sense, there must be a good life ahead, and my relation to that life must contain what prudentially matters. In these imagined cases, in each possible outcome, the life ahead would be very similar. If I could rationally be indifferent between these outcomes, my relation to both these lives must equally contain what matters. In such cases, this *Indifference Test* is a way of asking the prudential question.

It is also a way of asking whether identity is what matters. Of the possible lives ahead, only one would certainly be mine. If I could rationally be indifferent between these outcomes, this fact about my identity must be irrelevant. My relation to that future life, even if it would *not* be *my* life, must contain what matters.

One example of this kind is *Teletransportation*, in which my body would be destroyed and replicated. I claimed that, even if my Replica would not be me, my prospect would be as good as ordinary survival. My relation to my Replica would contain what prudentially matters, and, since my Replica's life would continue mine, my prospect would also contain what, in the desirability sense, matters. *Teletransportation*, even if it killed me, would be as good as conventional space-ship travel. Most of us find that hard to believe. Intuitively, this example provides a strong objection to my view.

Return, next, to *My Division*. To apply the Indifference Test, we must assume that, if I divided, the two resulting people's lives would be very like mine. I claimed that, even if I would be neither of these people, my relation to each of them would be about as good as ordinary survival. And, in this example, I gave an argument for that conclusion. Unlike the imagined case of *Teletransportation*, this example supports the view that identity is not what matters.

In some cases, the Indifference Test breaks down. In Bernard Williams's imagined case, which we can call the *Brain Zap*, my brain would be remodelled in a way that would destroy all psychological connections.⁷ The resulting person, though he would have my brain and body, would be psychologically quite unlike me. Suppose we believe that, as Williams argues, the resulting person would be me.

Even if we believed that we would survive such an operation, most of us would reject this prospect. We would strongly prefer to remain psychologically continuous with ourselves as we are now, and we would believe that preference to be rational. Ought we to conclude that such psychological continuity is part, or all, of what prudentially matters? That does not follow. Psychological continuity may instead be part of what makes survival good. Discontinuity, we may believe, would make our lives go worse. That could be enough to explain why we should reject such operations.

As this shows, when we consider such cases, we cannot use my Indifference Test as a way of asking the prudential question. For this test to work, we must compare outcomes in which the possible lives ahead would be very similar. That is how, in asking what we have reason to prefer, we are also asking whether our relation to these possible lives gives us reason for special concern. We might know that, in both of two possible outcomes, there would later be someone who would be just like us, and who would complete our book and care for our children, and we might ask whether it mattered whether that person would be us. If the possible lives ahead would *not* be very similar, what we can rationally prefer does not depend only on our relation to these lives. The desirability and prudential questions would not here coincide.

It may be objected that, if we add one assumption, we could still use the Indifference Test. If we chose to have our brains remodelled, in a way that gave us a wholly new psychology, our future would be very different. But we could suppose that, whether or not we had this operation, our future would be about *as good*. It may then seem that, if identity were what prudentially mattered, we ought to be indifferent which of these futures we would have. Both would be as good, and, in being *ours*, both would have the right relation to us. If instead we would have good reason *not* to be indifferent, these futures cannot both have the right relation to us.

This reasoning is, I think, unsound. One objection is that even if, with our new psychology, our future would be as good, the psychological discontinuity might make our life as a whole worse. This objection might be met if our new psychology would make our future better. Even with

the discontinuity, our life as a whole might then be as good.

To make the issues clearer, let us suppose that, with our new psychology, our life would as a whole go *better*. The loss of unity, we can suppose, would be outweighed by other gains. It might now be claimed that, to be rational, we should welcome this operation.⁸ But we might still disagree. Even if we believed that, with a new psychology, our life would go better, we might strongly prefer to keep our present character, desires, and other features. And we might believe that, in having that preference, we would *not* be less than fully rational.

On the reasoning suggested above, if we have that belief, we ought to conclude that identity is not what matters. If we can rationally reject this operation, even though it would make our lives go better, that must be because we would not have the right relation to our improved future. It cannot be enough that this future would be *ours*. What matters must be what is missing: the psychological continuity that, in ordinary lives, goes with identity. That must be, or at least be part of, what prudentially matters.

I believe, however, that we could still reject this reasoning. This argument assumes that, in such cases, we could not rationally prefer what would make our lives go worse. And that assumption can be challenged. If some event would make our lives go worse, that may always be a good reason to want this event not to occur. But such reasons may be outweighed. And what is relevant here is this. When we are comparing different ways in which our lives might go, we can rationally, I believe, give special weight to some of the features of our actual lives, and can therefore rationally prefer a future that would be somewhat worse for us.

Robert Adams gives a relevant example of a different kind. Suppose that I am happily married, and love my children. I must agree that if I had met a different partner, and had different children, my life might have gone even better. But I may be glad that my life did *not* go, in those ways, better. I may prefer to have my actual partner, and my actual children. Adams claims, I believe correctly, that such a preference would not be irrational. Judged from our present point of view, we can rationally prefer that our life be, on the whole, worse.⁹

Similar remarks apply to Williams's imagined operation. We are now supposing that, if brain-remodelling gave us a new psychology, our lives would as a whole go better. If we believe that it would be rational to reject this operation, we need not conclude that identity is not what prudentially matters. In such a case, what we can rationally prefer may diverge from

what would be best for us. If that is true, the Indifference Test would not here be an indirect way of asking the prudential question. We would have to ask directly whether, if we were about to have our brains remodelled, we would have as much reason to be specially concerned about the rest of our future.

To make that question vivid, we could suppose that, after this operation, we would be in pain. Would we have reason, Williams asks, to be specially concerned about that pain? Peter Unger often asks a similar question. After describing our relation to some imagined future person, Unger asks whether, to save that person from some ordeal, it would be rational for us, in a self-interested way, to choose to endure some lesser pain now.¹⁰ Unger's *Pain Avoidance Test* has one advantage. As Williams notes, our concern to avoid pain is almost unconnected to the distinctive features of our psychology. So we can apply this test even when our imagined case involves all kinds of psychological change.

My Indifference Test has, we have seen, a more restricted use. If we must sometimes directly ask the prudential question, why not do this all the time? Why not treat the Indifference Test, and the desirability question, as mere distractions?

Part of my answer would be this. Unger's *Pain Avoidance Test*, though useful, seems to me too narrow. Our desire to avoid pain is only one of the kinds of concern that we have about our future. This desire is both negative, and what Williams calls *conditional*: it is a desire that, *if* we survive, we do not suffer. We have other desires that are positive and categorical, since these desires could not be fulfilled unless we survive. Since such desires are not covered by Unger's Test, it is worth asking other questions.

We might switch to a positive version of the prudential question. Return to the case of brain-remodelling. We are now assuming that, if we chose to have this operation, our very different future would be worth living. Would we have reason for special concern about this future? Should we look forward, gladly, to what lies ahead? Should we be prepared to make sacrifices now, to make this future even better?

In trying to answer these questions, we may learn something. But in this example, since our psychology would be transformed, it is hard to focus on the prudential question. While our desire to avoid pain is almost unconnected to the rest of our psychology, that is not true of our desires for much of what makes our lives go well. Unless we are Narrow Hedonists, whether we look forward to our future will in part depend on whether that

future would fulfil some of our distinctive aims, projects, and ambitions. If we were about to lose all of these distinctive features, as would be true in Williams's example, our concern about our future would be knocked off balance. That would make it harder to decide what prudentially matters.

This question is easier to answer in cases that involve no such break in psychological continuity. And when we consider such cases, we should, I believe, use the Indifference Test when we can. Since this test compares two outcomes, it is more discriminating than the simple question whether some relation would contain what matters. Unlike Unger's Pain Avoidance Test, this test engages all of our concerns about our future. And, since this test compares outcomes that are very similar, it avoids engaging any other, irrelevant concerns.

3 Special Concern

Our question is about our reasons for *special* concern about our future. How should we understand such concern?

By 'special' we should not just mean specially strong. We may have other concerns that are just as strong. We might define this concern as self-interested, or as the concern we have for our own future. But that definition would make it trivially true that such concern can be only about ourselves, and that would regrettably restrict the scope of our discussion. It is worth asking whether we could have reason for such special concern about people other than ourselves. Should I have such concern, for example, about my Replica on Mars, or about the two resulting people in *My Division*? And we want these to be substantive questions, not ones whose answer is true by definition.

Let us therefore define this concern as the *kind* of concern that we have about our own future. When we define some attitude as the kind that we typically have toward some object, that definition allows that this attitude might have other objects. Thus, if we define parental love as the kind of love we have for our own children, it makes sense to suppose that we might have such love for some children who are not ours.

John Perry describes one kind of concern about our own future.¹¹ This he calls *derivative* concern, since it derives from our present desires, or projects. Perry points out that, if we have such projects, we shall probably have reason to want to remain alive and well. We are likely to be the 'best executors' of our own projects; and, if that is true, our own survival will

increase the chance that our projects will be fulfilled.

This kind of derivative concern is not restricted to our own well-being. As Perry notes, if we were about to be destroyed and replicated, we ought to have as much derivative concern for our Replica's well-being. This Replica would be just as able, and as likely, to fulfil our projects. Some of our projects, moreover, could be fulfilled by people who were not even similar to us.

As these remarks imply, Perry's derivative concern is very different from our ordinary concern about our own future. That is shown by Perry's claims about his attitude to future pain. If he knew that his Replica would be in great pain, Perry claims that he would have no reason for special concern, unless that pain would interfere with his Replica's fulfilment of his projects. If that pain would be, though intense, fairly brief, there might be no such interference. Perry concludes that, if he knew that *he* himself would later have to endure such pain, he would have no reason now for special concern. That is not what most of us believe, nor how most of us would feel.

Perry notes one way in which his derivative concern might be restricted to ourselves, and come closer to what most of us feel. Some of our desires or projects may essentially involve ourselves. We may want, not merely that some project be achieved, but that *we* achieve it. And we may have what Perry calls 'the ego project': we may simply want it to be true that we survive, and enjoy life. Even on Perry's account, we would then have reason to want it to be us who would survive. We are the only people who could fulfil *this* project. But, in having this project, we are *directly* concerned about our own future. Since Perry's account only covers derivative concern---concern which derives from our direct concerns---he says nothing about what such direct concern involves, or how it might be justified.

According to some other writers, our concern about ourselves is merely one instance of a wider phenomenon. Most of us are specially concerned about the well-being of certain people, such as our friends or relatives. Let us call this *affective special concern*. Why can we not, these writers, say, have such concern for ourselves?¹²

Affective concern covers more than derivative concern. It is direct, and it covers all aspects of well-being. But it still leaves something out. Perhaps some people care about themselves only in the kind of way in which they care about certain other people. But, when we think about our futures, most of us have another attitude, which might be called

primitive or instinctive concern. Such concern has, in Mark Johnston's words, a 'special and urgent quality'.¹³ This concern may not be *stronger* than our concern for certain other people, such as our own children. But it seems distinctive.

This distinctiveness seems embedded in some of our concepts. Thus it is only our own experiences to which we can look forward. And we can fear or dread only our own pain. Though I may care more about my child's pain, I cannot, it seems, fearfully *anticipate* that pain. We might call this *anticipatory* concern.

Anticipation seems, then, to presuppose identity. If I cannot anticipate my child's pain, that seems to be because this will not be my pain. If that is so, it may raise a problem for what, in my Chapter 14, I called the *Moderate View* about what matters. On this view, we have reason for special concern about any future person between whom, and ourselves now, there will be psychological continuity. And we have reason for such concern even if this future person will not be us. But if such concern includes anticipatory concern, and we cannot anticipate pains that will not be ours, this part of the Moderate View would be undermined.

Moderates might claim that, if anticipation does presuppose identity, that is merely true by definition. They might propose a wider concept, *quasi-anticipation*, which does not presuppose identity. Is this an acceptable reply? Is there a coherent and useful sense in which we could be claimed to quasi-anticipate someone else's pain?

This question seems to me difficult. We can first recall the corresponding question about memory. I am inclined to believe, with Shoemaker, that we could seem to quasi-remember past experiences, from the *inside*, without thereby assuming that it was we who had those experiences. In defending the coherence of such quasi-memories, I appealed to Williams's claim that we should not confuse imagining being someone else, such as Napoleon, with imagining *ourselves* being Napoleon. When we imagine being someone else, our own identity need not enter into what we imagine. This analogy may help to distinguish quasi-remembering having an experience from quasi-remembering ourselves having it. But, if we transfer this analogy to quasi-anticipation, the resulting concept may be too wide. The analogy suggests that, to quasi-anticipate having future experiences, it is enough to imagine these experiences from the inside: to imagine what these experiences would be like, for whoever had them. In that sense, we could quasi-anticipate the pain of a complete stranger.

When I say that this notion may be too wide, what I have in mind is this.

According to my Moderates, for me to have reason for quasi-anticipatory concern about the pain of some future person, that person must be psychologically continuous with me. My relation to that future person is, on the Moderates' view, special. If I could quasi-anticipate *anyone's* pain, this view would again be undermined. Moderates might reply that, though I *could* have such concern for anyone, I only have *reason* for such concern in the case of a future person who will be psychologically continuous with me. But that reply may not seem enough, since it does not explain what is distinctive in such anticipatory concern.

If that reply is not enough, Moderates might claim that, to be able to quasi-anticipate some future person's pain, we must believe that this person will be psychologically continuous with us. That does not seem plausible, as an account of our actual anticipatory concern. But Moderates might claim that, in this respect, our attitudes should change.

I do not know whether Moderates could defend such claims. Perhaps anticipatory concern presupposes a belief in identity, and in more than a trivial analytic way. If we believe that our Replica will not be us, there may be no coherent and relevant sense in which we could quasi-anticipate our Replica's pain. If that is so, this may threaten the Moderates' claim that psychological continuity gives us reason for anticipatory concern. But it would not threaten what I called the *Extremist View*, since Extremists deny that we ever have such a reason.

I have described two kinds of non-derivative special concern about our own future. We can now ask: Do we have reasons for these kinds of concern? If we do, what are these reasons, and how strong are they? If some future pain will be *ours*, is that a reason to be specially concerned about it? And, if so, would such concern merely be rational, or would it also be rationally required? Would it be irrational not to have such concern?

Besides an appeal to a purely instrumental desire-based theory about reasons, there are two other ways of rejecting these questions. It may first be claimed that, since we are bound to be specially concerned about ourselves, there is no point in asking whether such concern is rational, or well-grounded. Why question, or try to justify, what is inevitable?

Even in the case of attitudes that really are inevitable, this claim seems to me mistaken. If we could not conceivably lack some attitude, the question of its justification would take a different form. This question would not be

practical. But it might still be worth asking. And we might decide that, in having this attitude, we are, even if unavoidably, irrational. Perhaps, for example, we cannot avoid being biased towards the future. It is still worth asking whether this bias is irrational.

There is a different reply. Even when we are bound to have some attitude, or concern, its strength may vary. While it may be inconceivable that we could lack all concern for our own future, this concern can be outweighed by others, such as a concern for certain other people. That makes it a practical question how strong our reasons are for this concern.

It may next be said that, in asking for our reasons for such concern, we need not consider the metaphysics of personal identity. The relevant question is only whether having such concern is good for us. It might then be said that, since the answer to that question is clearly Yes, that is the only justification that we need.¹⁴

These claims are also, I believe, mistaken. First, it is not clear that this special kind of concern *is* good for us. Our lives would indeed go worse if we had no concern for our own future, and they would probably go worse if this concern was not much greater than our concern for any single stranger. But this concern need not be anticipatory concern. It could be a mixture of derivative and affective concern. If we had these other kinds of concern, without anticipatory concern, our lives might on balance go better.

Second, the relevant question is not only what would be best for us. Indeed, for that question to be relevant, we must have a non-pragmatic reason to care about our own well-being. Pragmatic justifications are quite different from, and cannot replace, justifications that appeal to direct, non-pragmatic grounds for concern. If I knew that I shall soon be cruelly tortured, it might be better for me if I did not care about my future. But that does not show that I would have no reason for concern.

In considering our reasons for special concern, we can ignore derivative concern. Of the two kinds of direct concern, we can start with affective concern. Adams suggests that such concern need not be justified. 'Love', he writes, 'need not be based on reasons. Few would suppose that our special concern for ourselves and our own future needs reasons.'¹⁵

It is controversial whether love, or friendship, needs reasons. Several writers would reject Adams's claim. They might agree that we do not fall in love, or have friends, *for* certain reasons. But, on their view, there have

to be features of our friends, or those we love, which justify that love. We can call this the *evaluative* view.

Our concern for ourselves may, in part, be justified in this evaluative way. We may care about ourselves because we like or admire the kind of person that we are. But this does not seem true of much self-concern. For one thing, while we can choose our friends, we cannot choose ourselves. Partly for that reason, our concern about ourselves may have little to do with our own qualities. While some kinds of self-loathing may destroy self-love, there are many people whose concern about their own well-being in no way depends on their admiring or liking themselves.

We are concerned about ourselves, not chiefly because of our qualities, but because we *are* ourselves. Our concern for others cannot take this form. Thus I cannot love *you* simply because you are you. Everything is identical with itself. And this property of self-identity cannot relate you to me. It does, however, relate *me* to me. Even if I cannot love you, regardless of your other properties, just because you are you, I may care about myself simply because I am me. Though everything is identical with itself, only I am me.

Numerical self-identity, or my being me, is not a qualitative matter. And the evaluative view fails to capture what is distinctive of anticipatory concern. Consider something narrower: our concern about our own present feelings. I am specially concerned about my present pain, not because I love or admire myself, but because this is what *I am now* feeling. I could have such special concern even if, while suffering from amnesia, I knew nothing about myself. Anticipatory concern seems to have the same pure quality. It is aroused simply by the thought that it will be me who is in pain. What I take myself to be like seems, here, irrelevant.

Return now to Adams's view about affective concern. On this view, love and friendship need no reasons. Such a view may apply to some of these attitudes. But, as before, it seems not to cover ordinary self-concern. We do not merely happen to care about ourselves, for no reason. We believe we have a reason for such concern, and we naturally take this reason to be the fact that we *are* ourselves.

Our love for our relatives may give Adams a better analogy. For one thing, just as we do not choose ourselves, we do not choose our genetic relatives. And our love for some of our relatives, especially, perhaps, our children, may be claimed not to be based on reasons. But such a claim takes 'reasons' to mean 'desirable or attractive features'. Perhaps we do not love our children because of what they are like. But we may love them

because they are *our* children, or because they are specially dependent on us. So this analogy cannot help to show that we need not justify, or do not have grounds for, self-concern.

Our concern about ourselves seems, then, to be different from our concern for other people, and to be grounded in a different way. It is not just a matter of our liking or admiring ourselves. And, unless we accept a purely desire-based or instrumental theory about reasons, we assume that we have reason for such concern.

That reason, most of us assume, is provided by the fact of personal identity. We care specially about our future because it will be *our* future. Is that assumption, I shall be asking, justified? Is identity what prudentially matters?

This question is closely related to certain others. Thus it is often claimed that the principles of justice are grounded in the separateness, or non-identity, of different persons. And it is claimed that, in applying distributive principles, we should take, as the important *units*, people's whole lives. These claims may assume that identity is what prudentially matters.

One connection comes through the idea of *compensation*. To explain the need for distributive principles, several writers claim that one person's burdens cannot be compensated by benefits to other people. That is why it is morally important to *whom* benefits and burdens come.

If we care about certain other people, benefits to them may give us one kind of compensation. But that would be merely what I have elsewhere called *quasi-compensation*.¹⁶ In the same way, if I care about my own future, I could be quasi-compensated now by the fact that benefits will later come to me. But what is morally more important is *pure* compensation. On the view that most of us accept, I could be compensated by future benefits even if I do *not* now care about my future. For me to receive such pure compensation, it is enough that these benefits will come to me.

On this view, pure compensation is made possible by personal identity. That is close to the view that identity is what prudentially matters, and that identity is what justifies anticipatory concern. So, if we decide that identity is not what matters, that may affect our beliefs about the scope and significance of compensation. We may be led to qualify the claim that, in applying distributive principles, we should take whole lives to be the relevant units. We might conclude that we should extend these principles

even within lives. Or suppose we decide that anticipatory concern is never justified. We might then conclude that there cannot be pure compensation over time. That would radically affect our distributive principles.

Personal identity, my book argued, is not what prudentially matters. Nor does it have moral significance. These claims I sometimes called part of a Reductionist view. But that was a mistake. Though these claims are part of my view, they are rejected by many Reductionists. We might believe that personal identity just consists in physical and/or psychological continuity, but still claim that identity is what matters. Many people hold this view.¹⁷

After arguing that identity is not what matters, I went on to claim that, if any relation matters in these ways, this relation is psychological continuity and connectedness, with any cause. But that is a separate claim, which needs a separate defence.

My book gave two arguments for the unimportance of identity. One appealed to a general claim about the implications of Reductionism. The other appealed to the special features of the imagined case in which I divide. I shall start with the general claim, which I presented in a quite inadequate way.

Part Two

4 Reductionism about Importance

I argued:

- (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.

Premise (1) is Reductionism; (2) we might call 'Reductionism about Importance'.

Mark Johnston rejects this argument.¹⁸ He calls it an *Argument from Below*, since it claims that, if one fact consists in certain others, it can only be those other *lower* level facts that are in themselves important. 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 have a derived significance. They matter, not in themselves, but because they constitute the higher level fact.

Should we accept either of these arguments, and, if so, which? In considering this question, it will be easier to start with subjects other than personal identity.

In some cases, Johnston and I would not deeply disagree. Suppose we accept that one of two facts consists in the other, and that it always matters whether these facts obtain. On my view, the first fact would matter only because it consists in the second. Johnston might hold the opposite view, on which the second fact would matter only because it constitutes the first. But, though these views conflict, there would be little need to choose between them.

In other cases, more is at stake. Suppose that, if

(a) someone is alive,

this fact consists in the fact that

(b) this person's heart is beating, and this person's other vital organs are functioning.

Since it often matters whether (b) obtains, and (a) consists in (b), I would agree that, in one sense, it often matters whether (a) obtains. It often matters whether someone is alive. But on my view, just as (a) consists in (b), (a)'s importance consists in, and derives from, (b)'s importance. (b), moreover, does not matter in itself. (b) is important only when, and only because, it is among the causes of the fact that

(c) this person thinks, and feels, and acts.

Only (c) matters in itself. And (b) has no importance when it is true that (b)'s obtaining cannot affect whether (c) obtains. That is true, for example, when damage to someone's brain makes this person irreversibly unconscious. On my view, if we were in that state, it would not matter that we were still alive.

In defending this view, I would appeal to an Argument from Below. Suppose we ask in advance how we would want to be treated if, after some accident, we would be in this state. If we would later still be alive, this fact would consist in the fact that our heart was still beating, and our other organs were still functioning. On my argument, we should ask whether these other facts would 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, as I believe, the answer is No, we should conclude that it would not matter that we were still alive.

If Johnston is 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 heart 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 heart 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.¹⁹

Consider next the moral question that such cases raise. Some people ask 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 people take a different view. These people might appeal to a Johnstonian Argument from Above, since 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 this argument has, I admit, some plausibility. The word 'kill' has such force that it can seem morally significant whether it applies.

Consider next another case involve killing. Suppose that *White* tricks you into drinking poison. Before this poison takes effect, *Grey*, knowing that you are about to die, intervenes in a way that kills you. We can suppose that *Grey* acts for either of two reasons. *Grey* might kill you as the inevitable side effect of saving someone else from serious injury. Or *Grey*'s aim might only be to spare you the agony that this poison causes. In either case, *Grey* knows that her act will not be worse for you.

If *White* hadn't poisoned you, you wouldn't have died. On a consequentialist broadening of these concepts, that is enough to make *White* count as harming and killing you. But suppose we appeal to the ordinary use of 'kill'. On that use, *White* doesn't kill you, because *Grey* intervenes and kills you instead.

If we are assessing *White*'s guilt, is it significant that *White* doesn't kill you? Kantians would answer No, since they believe that what matters is the harm that people intend. Compare two people who both attempt murder, but only one of whom succeeds, because the other's shot just misses. According to Kantians, in assessing blameworthiness, it is irrelevant that this second person does no harm, since attempted murder is as bad as murder. But, to assess Johnston's argument, let us suppose that we reject this view. Suppose we believe that, if someone kills, she is guilty of something worse than merely attempting murder. On this view, such a person should feel more remorse, and may deserve more punishment.

Note next that, in my example, *White* does not merely attempt murder. It is not merely true that *White* tried to kill you. *White* acted in a way that

was sufficient for your death, and that, in the circumstances, was also necessary. If White hadn't poisoned you, Grey wouldn't have intervened, and you wouldn't have died. Since that is true, White intentionally and maliciously caused your death. Even if we reject the Kantian view, we may believe that, for moral purposes, that is what matters. We may think it morally irrelevant that, on our ordinary use of 'kill', White didn't kill you.

In defending this view, we might appeal to an Argument from Below. Though White didn't kill you, this fact consists in the fact that Grey, knowing that you were about to die from White's poison, justifiably advanced your death. We may claim that, considered on its own, this abnormality in the causal route has no moral significance. It is still true that you died because White poisoned you. We may thus conclude that White should be classed, morally, not with attempted but with actual murderers.

On a Johnstonian Argument from Above, that conclusion is unjustified. We could be told: 'This difference in the causal chain may not, as such, have any moral importance. But since it is morally important whether an act is a killing, and this difference in the causal chain constitutes the fact that White didn't kill you, it has derivative importance. It means that what White did was *not* as bad as murder.'

As before, the Argument from Below seems to me more plausible.

Turn now to a different subject. Suppose that, after trying to decide whether and 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 these causes.

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 premises, we ought to be persuaded by these arguments.)

On my Argument from Below, if the fact that someone's choice was 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 this 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.

Johnston might object that, in choosing these examples, I am avoiding cases where it would be *his* argument that seemed more plausible. There are, I concede, such cases. As I have said, many people believe that it must matter whether, when someone is irreversibly unconscious, stopping this person's heart from beating would be an act of killing. And, when we turn to personal identity, most people's intuitions strongly favour Johnston's view.

Since there are conflicting intuitions, we should turn to the underlying question which this disagreement raises.

5 Realism about Importance

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, in relation to those other facts, the first fact is merely conceptual. Thus, if someone is irreversibly unconscious, but this person's heart is still beating, it is a conceptual fact that this person is still alive. And if one person deliberately causes another person's death, but in a way that runs through the will of a third person, it is a conceptual fact that the first person doesn't kill the second.

As before, when I call these facts conceptual, I don't mean that they are facts only about our concepts. That some unconscious 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 she 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', or the concepts that these words express.

When we turn to ask what matters, the central question is this. Suppose

we believe 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? Or suppose we believe that, when one person causes another's death, it cannot matter in itself whether the causation takes a certain form. Could this fact matter if, and because, it makes it correct to say that the first person did not kill the second? Or suppose we believe that it cannot matter, in itself, whether someone's choice was caused in one of two ways. Could this fact matter if, and because, it makes it correct to say that this person's choice was free? If we answered Yes to these questions, we would be treating language as more important than reality. We would be claiming that, even if some fact does not in itself matter, it may matter if and because this fact's obtaining allows a certain word to be applied.

Such a view, I believe, is irrational. 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 some concept applies is not a further difference in what happens, this conceptual fact cannot, I believe, be significant.

On Johnstonian Arguments from Above, even in such cases, conceptual facts retain their significance. It always matters whether someone can be correctly called alive, whether someone's act can be correctly called a killing, and whether someone's choice can be correctly called free. It is irrelevant if these facts just consist in other facts which are in themselves trivial. What matters is *that* these words apply, not *why* they apply.

This view seems to me to get things the wrong way round.

Conceptual facts, I concede, might have indirect importance. If I am irreversibly unconscious, for example, it might have legal consequences whether I am still alive. Perhaps, if my heart goes on beating until the end of the tax year, that would be better for my heirs. Facts about personal identity have many consequences of this kind. Suppose that, if I was Teletransported, my Replica would be denied my property. That would be one reason to prefer a conventional spaceship journey. But, when we ask whether some fact *in itself* matters, we should ignore such effects.

There are other ways in which conceptual facts may indirectly matter. For example, since laws must be framed in language, whether some act is illegal depends on whether some word applies to some act. If this word does apply, that may always give us some moral reason not to act in this way. In such cases, it may be irrelevant *why* this word applies. But the

fact that this word applies does not *itself* provide our reason not to act in this way. That reason comes from our obligation not to break the law. So such cases do not count against my view.

Conceptual facts, I have claimed, cannot rationally be thought to be *in themselves* important. That claim may be too strong. It may sometimes not be irrational to care, for its own sake, whether some conceptual fact obtains. Suppose, for example, that I am playing cards. Even when no money is at stake, I might want to be dealt some card, simply so that I shall win. That desire may seem irrational. Why should I want a random process to achieve a particular result, merely because it is conventionally deemed that, in that case, I shall have won? But this, after all, is merely a game, in which such groundless desires can be indulged. Similarly, as Unger says, we sometimes care whether certain conventions have been satisfied. Suppose I learn that, in walking from A to B, I have walked more than 100 miles. Or suppose I learn that the name 'Parfit' originally meant 'perfect'. Though these are conceptual facts, they might not irrationally give me some pleasure. In the same way, if I chose to be Teletransported, I might regret the fact that there would later be no one living who could be claimed to be me. That could be like regretting the fact that, after my death, there will be no one living who will have the name 'Parfit'. But it cannot be rational to give much weight to facts like these.

Johnston brings a second charge against Reductionism about Importance. 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 Reductionism about Importance, we must conclude that nothing has any importance. 'This is not a proof of Nihilism', Johnston writes, 'It is a *reductio ad absurdum*'.²⁰

This is an excellent objection, which shows the need to qualify my claims.

It may help to distinguish first between reductionism about what exists, and reductionism about facts. If we are physicalists, we might claim that, like everything else that exists, people just consist in certain arrangements of fundamental particles. It is a different and less plausible claim that all *facts* about people just consist in *facts* about these particles. In the sense of 'just consist' that I had in mind, that claim may not be true, even on a physicalist view.

We are Reductionists, I claimed, if we believe that our continued existence over time just consists in certain other facts. I meant facts about physical and/or psychological continuity. Hence my claim that, to defend a Reductionist view, we must be able to describe these continuities in a way that does not presuppose personal identity. Suppose that, as some writers argue, no such description can be given. A Reductionist might reply: 'That is no objection. *Every* fact just consists in facts about micro-particles.' But that would not be an adequate reply. To defend Reductionism about personal identity, we must be able to describe, in a clear and illuminating way, the lower level facts in which, we claim, facts about our identity consist. And this relation of *consisting in* must be of a different, and closer kind.

According to Analytical Reductionism, this closer relation is entailment. On this view, statements about personal identity mean the same as certain statements about physical and/or psychological continuity. That is not my view. I doubt whether such a claim is true. And I believe that Reductionists should aim, not for conceptual analysis, but for what Mackie calls 'factual analysis'. They should try to describe the facts given which we can make true claims about personal identity.

Though statements about personal identity do not, on my view, mean the same as statements about physical or psychological continuity, these statements are, in a different and looser way, conceptually connected. Hence my claim that, if we knew all the facts about both physical and psychological continuity, we would have all the empirical input that we need. If we knew about those other facts, understood the concept of a person, and had no false beliefs about the kind of entity that persons are, we would know, or should be able to work out, any facts that there might be about the existence and identity of persons. We should be able to work out such facts because these would not be, in relation to those other facts, independent or separately obtaining. In learning that such facts obtain, we could not learn any more about reality. All we could learn is how we use the concept of a person.

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. We need chemistry, biology, neurophysiology, psychology, and much else besides.

I shall now summarise my response to Johnston's charge of Nihilism. My

book claimed that, if one fact consists in certain others, it can only be those other facts which matter. As Johnston's objection shows, that claim must be qualified. My claim cannot use the word 'consist' in its broadest sense: the sense in which, on a physicalist view, everything may just consist in fundamental particles. I have in mind a relation which, though not entailment, is conceptual. I am concerned with cases where, relative to the facts at some lower level, some fact at a higher level is, in the sense that I have sketched, merely conceptual.

My argument does not assume that, whenever there are facts at different levels, it must be the lowest level facts which matter. That is clearly false. Nor does my argument assume that, when some entity consists in certain others, it must be the constituents which matter, not the entity as a whole. That is also false. What I assume is that merely conceptual facts cannot be rationally or morally important. What matters is reality, not how is described. Though this view is, in part, Reductionist, we can call it *Realism about Importance*.

6 The Bodily Criterion

Return now to the question of what gives us reason to be specially concerned about our future. Most of us believe that we should care about our future because it will be *our* future. I believe that what matters is not personal identity but certain other relations. To help us to decide between these views, we should consider cases in which identity and these other 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, like Williams and Thomson, believe that persons just are bodies. And it is the view of several of the people who identify persons with human beings.²¹ We can call this the *Bodily Criterion*.

Suppose that, in

Head Transfer, 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. On their view, 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. On their view, 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, in the sense that this question would have no answer. 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, this person's 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 the person who wakes up tomorrow would be me.

This fact may seem all important. After all, if this person would not be me, I shall have ceased to exist. But, if this person would not be me, this fact 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 would 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.

If we are Realists about Importance, we should now conclude that neither of these facts could matter greatly. Since it wouldn't be in itself important that my head would be grafted onto this body, and in relation to this fact it would be merely a conceptual 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 they would be heavily outweighed by the fact that, unlike me, the resulting person would not be paralysed.

Johnston disagrees. He believes that identity is what matters. And, if he accepted the Bodily Criterion, he would appeal to an Argument from Above. He might say:

It may not matter, in itself, that your head would be grafted onto the rest of this other body. But, if that were to happen, it would constitute your ceasing to exist. You have good reason to want to continue to exist. Since that is so, you have good reason to reject this operation.

For those who accept the Bodily Criterion, this reply may seem compelling. When it is applied to our own existence, Realism about Importance is hard to believe. But I believe that, here too, we should accept this view.

The main question, I have said, is the relative importance of language and the world. 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 was 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 this 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 wording of this question 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 travelling in some ferry boat this afternoon, 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, 'Shall I be in pain?' But that would be misleading, since it would suggest that I am asking what will happen.

In the case of the medical treatment, I don't know what conscious state I shall be in. There are different possibilities. In the case of being sea-sick, 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 treatment, I shall be in pain. And it matters whether, while I am in the boat, 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 could be successfully grafted onto my brother's headless body. I ask whether the resulting person would be me. Is this like the case of the medical treatment, or the ferry boat? Am I asking what would happen, or whether what I know would happen could be described in a certain way?

On my view, I am asking the second. I already know what would happen. There would be someone alive tomorrow who would have my head and the rest of my brother's body. It is a merely conceptual question whether that person would be me. And that is why, even if this person would not be me, that wouldn't matter. I should accept this operation.

These claims, as I have said, are hard to believe. Suppose that I am unsure which criterion of identity to accept. On the Bodily Criterion, the person who woke up tomorrow would be my brother. On the other criteria, this person would be me. It may seem that, if I didn't know which of these would be true, I couldn't know what was going to happen.

That, I believe, misunderstands what is at stake. Whichever criterion we accept, there are two undoubted facts. I would know that

(a) the person who wakes up tomorrow will have my head

and that

(b) this person will have the rest of my brother's body.

On the Bodily Criterion, (b) makes it true that

(c) this person will be my brother, and I shall be dead.

Though (b) and (c) would be different facts, they would be very closely related. (b) would not *cause* (c) to be true. These facts are too close for that to make sense. Rather, (b) would constitute (c). On this criterion, if the person waking up would have the rest of my brother's body, that would be *what it was*, in this case, for it to be my brother who woke up, and for me to be dead. Since (b) would be what it was for (c) to be true, (c) does not tell me more about what would happen. In claiming that I would be dead, we would merely redescribe the fact that my head would have been attached to the rest of my brother's body.

Similar remarks apply to the other criteria. According to the Brain Criterion, the fact that

(a) the person who wakes up would have my head

makes it true that

(d) it would be I who wakes up, and my brother would be dead.

As before, (a) would not cause (d) to be true. Rather, (a) would constitute (d). On this view, the fact that this person would have my head would be *what it was* for this person to be me. In claiming that it would be I who woke up, we do not give further information about what will happen. We merely redescribe the fact that this person would have my head.

(c) and (d) cannot both be true, since I cannot be both alive and dead. That may suggest that these are different possibilities, either of which might be true. But, in a case like this, (c) and (d) would *not* be, at the level of what happens, different possibilities. Return to the case of someone who is irreversibly unconscious. Let us make one change in this example. Suppose that, since there is damage to this person's lower brain, this person's heart continues to beat only because it is attached to some machine. It is now a harder question whether this person is alive. As before, he cannot be both alive and dead. But these are not, here, different possibilities, either of which might be true. We are only trying to decide how to make our concept of death more precise.

In our imagined *Head Transfer*, we have a different question. The person who woke up tomorrow would clearly be alive. What this person could

not be is both me, and my brother. But, in trying to decide who this person would be, we are not comparing two hypotheses: two predictions about what would happen. On one view, in such a case, personal identity would be indeterminate. If that is so, we are merely trying to decide how to make our concept of a person more precise. We are merely choosing one of two descriptions.

Other people believe that, in such a case, there is a determinate answer to our question. These people might, for example, claim that given our actual concepts of a person, and of a human being, we are already committed to one of these two redescriptions. If that is so, this description could be claimed to state a fact, or truth, about what would happen. Thus, if the Bodily Criterion is correct, it would be true that

(c) the person who wakes up will be my brother, and I shall be dead.

It would then be false that

(d) it will be I who wakes up.

If the Brain Criterion is correct, (d) would be true, and (c) false. But, as before, these are not rival *predictions*.

There is another point. If we accept the Bodily Criterion, our claim should not be that, if my head were grafted onto the rest of my brother's body, that would cause me to die. Our claim is about what would constitute my dying. On our view, I would be dead if either

(e) my whole body has ceased to function,

or

(f) most of my body has ceased to function, though my head has been successfully grafted onto the rest of someone else's body.

Since (e) is the ordinary form of death, we naturally assume that death is bad. We assume that, if some kind of surgery would kill us, that is a reason to turn it down. But, if (f) is a different way of being dead, it would be irrational to assume that (f) must have (e)'s significance.

In the case of many concepts, it is not enough to know *that* they apply. We must know *why* they apply. One example is close at hand. There are different ways of being alive: the ordinary, valuable way, and the way that involves being irreversibly unconscious. It would be irrational to think that, because these are both ways of being alive, they must have the same

significance.

The same applies to (e) and (f). It is one thing for my whole body to be destroyed, or cease to function. It is quite another thing for my head to be successfully transplanted, and for my brain to continue to function. On the Bodily Criterion, these could both be ways of being dead. But I should not assume that, because the first is bad, the second must also be bad. Just as there may be nothing good in being alive, if I am irreversibly unconscious, there may be nothing bad in being dead, if my brain continues to function, as the brain of a conscious living person.

It is natural to assume that, if I shall not wake up tomorrow, that is what matters. But it is not enough to know *that* no one alive tomorrow will be me. We must ask *why* this is true, or what it is that would make this true. If someone will wake up, and live a life that seems to continue mine, what would it *be* for that person not to be me? When we see what this fact would involve, in this very unusual case, we may see that this way of dying should not matter to me.

7 The Narrow Psychological Criterion

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 some excellent philosophers who accept the Bodily Criterion, I am not cheating. It is worth trying to show these people that identity is not what matters. But I agree that we should reject this 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.²² Consider next those who believe that persons are animals of a certain kind, viz. human beings. As I have argued, we could take this view, but reject the Bodily Criterion.²³ 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, or with which it often coincides. But, when we consider cases in which personal identity coincides with these other facts, it is harder to decide whether identity is what matters. So, if we reject the Bodily Criterion, we should consider other cases.

Suppose that we accept the Narrow, Brain-Based version of the Psychological Criterion. On this view, if there will be some 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.

Suppose, 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 Narrow Psychological 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 we regard these as problem cases, ones in which we do not know what would happen to me. Return to my imagined *Physical Spectrum*. In each of the cases in this range, some proportion of my cells will be replaced, in a single operation, with exact duplicates. With some proportions--20%, say, or 50%, or 70%---most of us would be uncertain whether the resulting person would be me. ²⁴

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 an answer, we could know what is going to happen. And this conceptual question is not even 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 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's 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 this person will be in pain, can I expect to feel that pain? But these remarks don't help. These are just other ways of asking whether this person will or will not be me. In appealing to what I can expect, we do not explain what would make these different possibilities.

This difference, we may think, needs no explanation. It may seem enough to say: Perhaps this 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 different 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 me-ness, which either might or might not be there.

If we turn to the conceptual level, there *are* different possibilities. Perhaps that future person could be correctly called me. Perhaps he could be correctly called 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?

8 The Division Argument

I am now assuming that we accept the Narrow 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.

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 possible operations. In the first operation, 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 whole 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*. In this operation, both halves of my brain would be transplanted, into different bodies that were just like mine. Two people would wake up, each of whom has half my brain, and is, both physically and psychologically, just like me.

I argued:

- (1) In the *Single Case*, the resulting person would be me.
- (2) My relation to myself in the future would here contain what prudentially matters.
- (3) Whether this relation contains what matters must depend only on its intrinsic features.
- (4) My relation to myself tomorrow, in the *Single Case*, is intrinsically the same as my relation, in *My Division*, to each of the two resulting people.

Therefore

- (5) My relation to each of these people must contain what matters.

However

(6) It is not true that each of these people would be me.

Therefore

(7) Personal identity would not, here, be what matters.

This we can call *the Division Argument*. Should we accept this argument?

Of those who reject premise (1), some appeal to the Bodily Criterion. On their view, for some future person to be me, he must have my body. But even for those who accept this view, this objection can be met. For some future person to be me, he need not have all my body. Even if half my brain and body were destroyed, the remaining half might, after reconstructive surgery, continue to function as the brain and body of a living person. Even on the Bodily Criterion, it must be admitted that this person would be me. So, if we accept this criterion, we must claim that, for some future person to be me, he must have *enough* of my body. And we must allow that half might be enough.

My argument could now appeal to a different pair of cases. In the *Single Case*, which we have just described, I would survive with only half a body. In the *Double Case*, my whole body would be divided, and each half would continue to function as the body of a living person. With this change in the example, those who accept the Bodily Criterion must accept my argument's first premise. The rest of the argument would be unaffected.

Some people would give a different reason for rejecting premise (1). Any criterion of identity, these people believe, must meet a certain condition. According to what we can call

(8) or *the Intrinsicness Requirement*: If some future person will be me, this fact must depend only on the *intrinsic* features of the relation between me now and that future person. It cannot depend on whether this same relation also holds between me now and some other future person.

My argument violates this requirement. In both the *Single* and the *Double Case*, someone will wake up with the left half of my brain. We can refer to this person as the *Left-Brained Person*, using this phrase, not as a name, but as a description. Whether the Left-Brained Person will be me depends, my argument assumes, on what happens elsewhere to the right half of my brain. That is not an intrinsic difference in my relation to the Left-Brained Person.

Though my argument rejects (8), its third premise could be more fully

stated as follows:

(3) or *the Intrinsicness Principle*: If my relation to some future person contains what matters, this fact must depend only on the intrinsic features of the relation between me now and that future person. It cannot depend on whether the same relation holds between me now and some other future person.

It may seem that, given (3)'s similarity to (8), we should either accept both claims, or reject both. But that, I believe, is not so. (8) is a claim about identity, and (3) is a claim about what matters. If personal identity were what mattered, these two claims would be tied together. But, if my argument is sound, identity is not what matters. So, to assess my argument, we must consider these claims separately, and judge each on its merits.

As I have already argued, we should reject (8). No plausible criterion can meet this requirement. (8) implies that, if half my brain and body were destroyed, I would cease to exist. That would be so even if, with reconstructive surgery, what was left of my brain and body continued to function in a normal way. When we imagine such a case, that conclusion is very hard to believe. Nor are there arguments for (8) which force us to that conclusion.

I have yet to defend (3). But, as we shall see, (3) raises different questions. Our reasons for accepting (3) do not conflict with our reasons for rejecting (8). The similarity between these claims is not, though, an accident. The Division Argument might be summed up as follows:

On any plausible criterion, personal identity could depend on extrinsic facts. Since what matters does not depend on such facts, identity cannot be what matters.

My argument has three other premises. According to (2), in the *Single Case*, my relation to myself tomorrow would contain what mattered. This premise is not controversial. Many would accept (2) because they believe that identity is what matters. I accept this premise for a different reason. My argument leaves it open what the facts or relations are that, in ordinary cases, give me my reasons for special concern about my future. But, whatever those facts or relations are, they would be present in the *Single Case*. The person who wakes up would be psychologically continuous with me, because he would have half my brain. In my revised version of this case, he would have half my body. He would also be me. Given those facts, I assume, I would have my ordinary reasons for special

concern. Those reasons would not be removed, or reduced, by the fact that, before waking up, I would have lost other parts of my brain and body.

Premise (4) is also not controversial. (4) merely states that, in both the *Single* and the *Double Case*, there would be no intrinsic difference in my relation to the Left-Brained Person. These cases differ only in what happens, elsewhere, to the right half of my brain and body.

According to (6), my remaining premise, I would not, in the *Double Case*, be each of the resulting people. In making that claim, I assume that, before *My Division*, I am a single person, and that, after *My Division*, there would be *two* resulting people. Both assumptions can be questioned. But they are hard to deny. And, on these assumptions, premise (6) is uncontroversial. If I am a single person, and there would be two resulting people, it cannot be true that each of them is me. That would be a contradiction. If each of them was me, they would each be one and the same person: me. So they couldn't be two different people.

If we accept these six premises, we ought, I believe, to accept this argument's conclusion. (1) to (4) entail (5), and (5) and (6) together entail (7). In a case like *My Division*, personal identity would not be what matters.

In his discussion of this argument, Johnston starts by considering my argument's third premise, which I have restated as

(3) or *the Intrinsicness Principle*: If my relation to some future person contains what matters, this fact must depend only on the intrinsic features of the relation between me now and that future person. It cannot depend on whether the same relation holds between me now and some other future person.

Johnston suggests that, if we find (3) plausible, that is because we ignore the possibility of a case like *My Division*. In his words, 'we suppose that we will have at most one future continuer, hence that whether some process represents our continuation only depends upon intrinsic features of that process, and hence that whether some process is rightfully taken to ground direct future concern can only depend upon intrinsic features of that process.'²⁵

This diagnosis seems to me far-fetched. (3)'s plausibility does not, I think,

result from Johnston's rather complicated line of thought. There are countless relations whose significance we take to depend only on their intrinsic features. (3) merely claims that we are now considering another such relation. And, if we find (3) plausible, that is not because we ignore the possibility of a case like *My Division*. (3) seems plausible when, and because, we are considering this imagined case. Suppose that we accept my argument's first two premises. We believe that, in the *Single Case*, the Left-Brained Person would be me. And my relation to this person, i.e. to myself tomorrow, would, we believe, contain what matters. We may then believe that, even if what happens is the *Double Case*, my relation to the Left-Brained Person must still contain what matters. As I wrote, since this relation is intrinsically the same, '*Nothing is missing*'. That is the thought which can make (3) seem compelling.

That thought may seem a mistake. After all, in the *Double Case*, the Left-Brained Person cannot be claimed to be me. And, if he *isn't* me, surely something *is* missing. *I'm* missing. 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 difference is only that, in this *Double Case*, I also have the very same relation to the other resulting person. When this relation holds between me now and a single person in the future, that person can be called me. If this relation held between me now and *two* future people, I could not be called one and the same as each of these people. But that would not be a difference in the nature of this relation. In the *Single Case*, where half my brain would be unharmed, my prospect is survival. That prospect contains what matters. In the *Double Case*, where both halves would be unharmed, nothing disappears.

Johnston also questions (3), which I am now calling the *Intrinsicness Principle*. Johnston points out that, in the case of certain facts or relations, we take extrinsic features to be relevant. His examples, drawn from Ernest Sosa, are 'exclusive ownership, winning, unique achievement, and intimacy.'²⁶ In such cases, the importance of some relation may depend on its uniqueness, or the absence of competitors. Why could not the same be true, Johnston asks, of the relation that matters in our survival?

Of Sosa's examples, the best is intimacy: the relation that is involved in romantic, conjugal, or sexual love. It is sometimes claimed that, if two people have such love for each other, all that matters is the quality of their relation. On this view, so long as that relation is unaffected, it doesn't matter if either of these people also has such love for someone else. But most people take a different view. These people believe that, for such

mutual love to have its full value, it should be exclusive.

If we find this second view plausible, does this cast doubt on (3), the Intrinsicness Principle? I believe not.

We should remember first that, in asking what prudentially matters, we are not asking about the *value* of the relation we might have to some future person. By 'What matters in survival?', we do not mean 'What makes our survival good?' We mean: 'What makes our survival matter? Why should we be specially concerned about our future?' But if we add one assumption, we can bring these questions together. We can suppose that, if I divide, the two resulting people's lives would be very like the life that, if I had not divided, I would have had. On that assumption, the desirability and prudential questions coincide. We can ask whether, if I was about to divide, my prospect would be as good as ordinary survival.

Return now to Sosa's analogy. Suppose that, to have its full significance, mutual love should be exclusive, or take a one-one form. The same might be true, Johnston suggests, of what prudentially matters. If that were so, and I was about to divide, my relation to the resulting people could not fully contain what matters. My prospect would be less good than ordinary survival.

Sosa's analogy seems weak. Our relation to ourselves in the future seems in many ways unlike mutual love. Perhaps Isolde's love for Tristan would have less value, even if it were unaffected, were she also to love King Mark. But nothing comparable seems true of my relation to myself in the future. If I choose to divide, I would not betray myself, or either of the two resulting people. As Sosa himself writes: 'We can see the sorts of value that would be threatened by having too many spouses; but I for one have no inkling of what important values would be endangered by [division]. . . it is not easy to see what values would be endangered in such cases *except* only for the true survival of the mainstream protagonist.'

This last claim is too sweeping. As several writers point out, division might not preserve all that is good in ordinary survival. Some important values might be endangered. One such value may be, precisely, intimacy. Sosa concedes that we could have too many spouses. But, if our single spouse divided, that might become true. If there are now two Tristans, what should Isolde feel and do? And, if Isolde has too many spouses, these Tristans have too few. It might help if Isolde also divided, since that would allow new couples to form. But it would be arbitrary which Tristan loved, and was loved by, which Isolde. And each new pair of Tristans and Isoldes would then have to share their world with the other pair. These

might both be changes for the worse.

There might be other bad effects. As Susan Wolf remarks, if I divided, each resulting person could inherit only half my property; and perhaps only one of these people could continue my career.²⁷ There might also be advantages. Perhaps the other resulting person could continue some career which I had reluctantly abandoned. But it might be claimed that, in most cases, the losses would outweigh the gains. That might make it true that, for most people, the prospect of Division would be worse than ordinary survival.

As we have seen, even if that claim were true, it would be irrelevant. We are asking what prudentially matters, or what gives us reason for special concern. If I shall soon be paralyzed, my prospect may be bad, but that does not show that I have less reason to care about my future. In the same way, even if Division would have bad effects, that cannot show that I have less reason to care about the futures of the two resulting people.

To improve our question, therefore, we should suppose that these people's lives would *not* be worse than mine. And we should suppose that these lives would go much as mine would go, if I did not divide. Though each of these people might be affected by the existence of the other, it is easy to imagine that, because of various facts about me, these effects would not be great, and that such good and bad effects would be roughly balanced. This is the only version of the case in which it is relevant to ask whether my prospect would be as good as ordinary survival. If each of these people would have lives like mine, but my prospect would be worse than ordinary survival, that must be because my relation to each of these people would not contain what matters.

Remember next that, in considering this case, our present question is only about the Intrinsicness Principle. There is of course one ground for claiming that my prospect must be worse than ordinary survival. I shall *not* survive. There will be no one living who will be me. But, in appealing to that claim, we would be assuming that identity is what matters. And, in considering my argument, we cannot appeal to that assumption. If my argument is sound, identity is not what matters. To ask whether this argument is sound, we must ask whether it is valid, and has true premises.

We are now considering one of these premises. According to the Intrinsicness Principle, if my relation to some future person gives me reasons for prudential concern, that must be because of the nature, or content, of this relation. It cannot depend on whether I also stand in the

same relation to some other future person. Should we accept this claim?

It may help to consider another analogy. Imagine a community of persons who are like us, but with two exceptions. First, because of facts about their reproductive system, each couple in this community has only two children, who are always twins. Second, because of special features of these people's psychology, it is of great importance for the development of each child that it should not, through the early death of its sibling, become an only child. Such children suffer psychological damage. People in this community therefore believe 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 in groups of three. 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 these children come in threes, they 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's having at least one sibling.

In the same way, I argued, 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 will have enough of my brain. When there will be only one such person, he can be described as me. If there would be two such people, we could not claim that each of these people is me. But that is as trivial as the fact that, if I had two identical siblings, they could not each be called my twin.

This analogy may be questioned. In my imagined community, people who start life as an only child suffer psychological damage. If that is the ground for believing that each child should have a twin, it would be easy to show, once triplets start to be born, that this belief is mistaken. We could appeal to the fact that no triplet suffers such damage. In the imagined case of *My Division*, there is no corresponding fact to which we could appeal. If we reject the Intrinsicness Principle, our ground for doing so would not be that, if psychological continuity took a branching or one-many form, that would have bad effects. Though there would be cases in which that would be true, such effects, I have argued, are irrelevant. In rejecting the Intrinsicness Principle, we would be making the quite different claim that, for psychological continuity to give us reason for special concern, it must take a one-one form.

To make the analogy closer, it could be revised. Suppose that, in my imagined community, only children do not suffer psychological damage. There is a different reason why, in this community, it is thought of great importance that each child should have a surviving twin. Love between twins, these people believe, enriches the lives of each. In having such value, it is like other kinds of love, such as that between parents and children, and that between sexual partners. But it has its own distinctive and valuable features.

Suppose next that, in this community, people believe that, for love between sexual partners to have its full value, it should be exclusive. But they do not have this belief about the love either of parents for their children, or of children for their parents. In most cases, each parent has two children, and each child has two parents. This does not reduce the value of these relations. Love for one's mother is not cheapened if one also loves one's father; nor is love for one's daughter cheapened if one also loves one's son. We can suppose that, in the case of certain families, there is only one remaining parent, and only one child. Some of these parents might claim that, because their relation to their child has the exclusiveness of romantic love, that gives it greater value. But most of my imagined people would rightly reject that view. Love of one's child, they believe, should *not* be such as to take this exclusive form.

What should these people believe about love between twins? Should such love be exclusive, or one-one? Someone might answer: 'Yes. If such love was had for more than one other person, it couldn't *be* love between *twins*.' But that answer would be superficial. At the start of my story, since children are always born as twins, my imagined people do not have, in addition to the concept of a twin, the wider concept of a sibling. But, when children start to be born as triplets, they would develop that wider concept. And they would then face a question which could not be settled merely by appealing to the meaning of the word 'twin'. To have its full value, should love between *siblings* take, like romantic love, a one-one form? Or is it like parental or filial love, which retains its value even when this love is had for more than one other person?

The second answer is, I assume, correct. And this version of my story now provides a closer analogy. The question I have asked cannot be settled by an appeal to psychological effects. It is a purely evaluative question.

The analogy is still not close. In asking what prudentially matters, we are not asking about the value of some kind of love. But this analogy may help to clarify our question. When some relation always in fact takes a one-one form, we may describe this relation in a way which makes this fact

a necessary truth. If this relation *could* take a one-many form, we should not be misled by that necessity. We would then face a substantive question. For this relation to have its special significance, must it take an exclusive, one-one form? Or does this significance depend only on this relation's intrinsic features: features that would be retained even if this relation took a one-many form?

Most significant relations are of this second kind. And, according to my Intrinsicness Principle, that is true of the relation that gives us reason for special concern. If we reject this principle, we must explain why, in the case of this relation, exclusiveness matters. In the case of sexual or romantic love, plausible explanations can be given. But in the prudential case, as Sosa claims, it is hard to see what explanation there could be. If there will be someone in the future who will be psychologically just like me, because he will have enough of my brain, I have a special relation to that person. That is the relation in which I stand to myself tomorrow. How is that relation altered, and made less significant, if I also have this same relation to a *second* future person? We cannot appeal here to such things as the loss of intimacy, or betrayal.

As before, there is one fact to which it is natural to appeal. If I had this relation to a *pair* of future people, they could not each be *me*. But this may be only like the fact that, if I have two siblings, they could not both be my twins. More important, as I have argued, this fact cannot be a ground for rejecting the Intrinsicness Principle. If that principle is correct, it is irrelevant that these people cannot each be me. So we must try to decide whether there are other reasons to reject this principle. There are, I believe, no such reasons.

Though Johnston questions the Intrinsicness Principle, his attitude to Division is, in the end, much like mine. It would be reasonable, he claims, 'to care about each of one's future fission products as if each were oneself'.²⁸ And the prospect of Division seems to him about as good as ordinary survival. He would pay no great price, he says, to ensure that only one 'product' survived, so that *he* would be that product, and would therefore still exist. With these remarks, Johnston seems to accept my argument's conclusion. He seems to agree that, in this kind of case, personal identity would not be what matters.

Johnston questions, however, the significance of this conclusion.

PART THREE

9 Johnston's Quarantining Manoeuvre

Johnston writes:

[T]here is a false apparatus of generalization at the heart of Parfit's argument against identity-based concern. If the Division case showed that a presupposition of such concern was always violated then there would be general consequences for our concerns. So if the existence of a superlative further fact were such a presupposition the general consequences would threaten. What is evidently not there in the Division case, the superlative further fact, is never there. However, since the relevant presupposition of self-concern is the holding of the determinate, ordinary fact of personal identity or difference, the case has no effect beyond the imaginative fringe. What is not there in the Division case is almost always there. Identity is still almost always what matters.²⁹

These remarks suggest that, according to my Division Argument, our 'identity-based concern' presupposes the existence of some 'superlative', or Non-Reductionist, 'further fact'. That is not so. The relevant premises were that, in the *Single Case*, the Left-Brained Person would be me, and that my relation to that person would give me my ordinary reason for prudential concern. Those premises make no appeal to any superlative further fact.

Johnston also claims that, even if my argument is sound, I should not have drawn a general conclusion. I went on to claim

(9) Since identity would not, here, be what matters, it is never what matters.

Johnston rejects (9). Though he concedes that, in this example, identity may not be what matters, he believes that we can *quarantine* this conclusion. We can claim that, in ordinary, cases identity is still what matters.

(9), I agree, needs more defence. In giving this defence, I shall start with a simplifying assumption. My argument implied that

(5) In *My Division*, my relation to each of the resulting people must contain what matters.

My relation to these people we can here call *M*.³⁰ This relation has two main elements. Each of these people would be psychologically continuous with me, and this continuity would have its normal cause: the continued existence of enough of my brain. (5) uses the word *contain* because my argument leaves it open which elements of *M* would matter. But, for the time being, let us ignore this complication. Let us take (5) to claim that, in this example, *M* would *be* what matters.

On this simplifying assumption, I would defend (9) as follows:

(A) In ordinary cases, if I would be *M*-related to some future person, that person would be me. In these cases, *M* and personal identity coincide.

(B) In *My Division*, *M* and identity would diverge. Though I would be *M*-related to each of the resulting people, it would not be true that each of them is me.

(C) My relation to each of these people would be what mattered.

(D) If there were other kinds of case in which *M* and identity diverged, these would also be cases in which it would be *M* that mattered.

Therefore

(E) Whenever *M* and identity diverged, it would be *M* that mattered.

Therefore

(F) Even when these relations coincide, it is *M* that matters. Identity is never what matters.

This we can call my *Divergence Argument*.

Of this argument's premises, (A) is uncontroversial, and the Division Argument supports (B) and (C). I shall return to (D).

Johnston seems prepared to allow (E), even if only for the purposes of argument. But he challenges my inference from (E) to (F). The natural view, as we both agree, is that identity is what matters. While my argument may show that we should revise this view, it cannot, Johnston claims, give us sufficient reason to reject this view. 'Reasonable adjustments are those in accord with what Quine calls the maxim of minimum mutilation.'³¹ (F) is a 'much more radical' revision than my argument requires. It would be more reasonable to stop at (E). On this 'locally modified' view, identity would 'almost always' be what matters. Only in a few cases would that not be true.

On Johnston's reading of Quine's maxim, when we acquire a firm belief which conflicts with some existing theory, we should try, in accounting for that belief, to make the smallest possible revision to our theory. That may often be a reasonable response. But there are cases in which it goes astray. Thus, if our theory was Ptolemaic astronomy, we should not simply add more epicycles. In some cases, we need a new theory.

We are not concerned here with a theory: only with the question of what, if anything, prudentially matters. But if we decide that, whenever personal identity and M diverge, identity would not be what matters, I believe that we *can* conclude that identity is never what matters.

First, this inference is of a kind that we often make, and find plausible. Suppose that X and Y---whether these be facts, or elements, or features---nearly always go together. We are convinced that X-and-Y matter, but we are not sure why. We ask what, more precisely, matters? There are several possibilities. Perhaps what matters is only X, or only Y. Or perhaps both matter, and each would matter on its own. Or perhaps what matters is essentially the conjunction, since neither would matter on its own.

To choose between these possibilities, the obvious test is to think about cases in which X and Y come apart. Suppose we decide that, in such cases, it is Y which matters. And we decide that, even when Y is on its own, Y has all of the importance that, in ordinary cases, is had by X-and-Y. That would strongly suggest that Y is what really matters. As Johnston himself writes: 'Such . . . thought experiments might . . . teach us something about the relative importance of things that invariably go together. Something we value non-derivatively might be shown to be a mere concomitant of what is really important.'³²

This is the reasoning that my argument employs. In all ordinary cases, personal identity and M coincide. When they diverge, M is what matters.

This strongly suggests that, in all cases, M is what matters.

Johnston rejects this reasoning. On his view, our conclusion should be *quarantined* to this imagined case, and other such special cases. Even if M would be matters when these two relations come apart, it is identity that matters when they coincide.

Johnston's view seems to me implausible. If we believe that, when two facts come apart, one of them is what matters, why should we think that the *other* is what matters when they coincide? Consider, for example, being alive, and being (often) conscious. If brain damage made us irreversibly unconscious, these two would come apart. Suppose we decide that, in such cases, there is no value in merely being alive. We decide that, without consciousness, life is worth nothing. It would then be strange to claim that, in ordinary cases, in which we have both life and consciousness, what matters is not being conscious but being alive.

I have claimed so far that, when two things almost always go together, we can reasonably judge their relative importance by asking what we should believe when these things come apart. Our conclusion could then plausibly extend to all cases. There are other grounds for thinking that, in this particular discussion, our conclusion can be general.

First, when people believe that personal identity is what matters, what they believe is that identity *must always* be what matters. On their view, if I shall soon cease to exist, my relation to someone else---someone who will not be me---cannot be as good as ordinary survival. If I believed that this relation would be as good, I must be making a mistake. This view cannot tolerate local exceptions. If we can describe a case in which my relation to someone else would have the significance of ordinary survival, this view must be abandoned. If identity could sometimes fail to be what matters, it cannot have the kind of significance that we intuitively think it has: a significance that must, we think, be universal.

Second, there are metaphysical grounds for drawing this general conclusion. It makes a difference how and why identity could fail to be what matters. We are not considering two separate relations, which generally go together, but might sometimes come apart. Rather, we have one relation which often constitutes the other. When M takes a one-one form, we can speak of identity; when M takes a branching form, we cannot. If we decide that, in a case of branching, M would have all of the importance that we normally ascribe to identity, it seems reasonable to conclude that, even when M does not branch, this relation is, or contains, what matters.

10 Johnston's Annexation Strategy

Johnston's quarantining manoeuvre does not, I have claimed, succeed. If we decide that, when M and identity diverge, M would be what matters, we can reasonably extend that conclusion to all other cases.

At certain points, Johnston suggests a different and bolder reply. In my imagined cases, he suggests, M would not matter in its own right. M's importance 'would be parasitic on the importance of identity'.³³ If that were true, my appeal to these cases would show nothing. If M's importance would be derivative from the importance of identity, we could claim that, even in these cases, identity would be what mattered.

On this reply, rather than admitting that these cases count against our view, and trying to quarantine that conclusion, we *annex* these cases, claiming to include them in our view. So I shall call this Johnston's *annexation strategy*.

In considering this reply, we should distinguish two kinds of case. Suppose we believe that, in some imagined case, our relation to some future person would contain what mattered. There are two ways in which this relation might diverge from identity. It might be indeterminate whether this future person would be us, or be determinately true that this person would *not* be us.

Even in cases of determinate non-identity, we might appeal to the annexation strategy. Suppose that, when we consider *My Division*, we are convinced that

- (1) neither of the resulting people would be me.

Even if we admit that

- (2) my relation to these people would contain what matters,

we might claim that

- (3) this relation's importance would be parasitic on the importance of personal identity.

In defending (3), we might use an Argument from Above. We might claim that

- (4) if my relation to the resulting people would contain what

matters, that is because in ordinary cases, in which this relation takes a one-one form, it constitutes personal identity.

I shall return to this claim. But it faces an obvious objection. If some relation is important only because, in ordinary cases, it constitutes identity, how can it have the *same* importance even when it does *not* constitute identity? If identity is what matters, how can it make no difference that the resulting people would *not* be me?

Johnston would reject (4), and for the reason I have just given. Johnston believes that, if some future person would determinately *not* be us, our relation to that person cannot fully contain what matters. If this relation did contain what matters, such a case, Johnston agrees, would count against the view that identity is what matters. But when he suggests the annexation strategy, Johnston is discussing cases in which personal identity would be indeterminate. *My Division*, he believes, is such a case. On his view, if I were about to divide, there would be no answer to the question of what would happen to me.

In cases of indeterminacy, Johnston's strategy is more plausible. Suppose that, when we imagine some case, we believe both that it would be indeterminate whether some future person would be us, and that our relation to this person would contain what matters. Though it is not true that this person would be us, it is also not true, Johnston says, that this person would be *someone else*---someone who is *not* identical to us. Our relation to this person must contain features which prevent this from 'being a case of simple non-identity'. That is why, though our relation to this person would not be identity, its importance 'would be parasitic on the importance of identity'.

On Johnston's annexation strategy, since this relation derives its importance from identity, we can reasonably regard this relation *as if it were* identity. In deciding that this relation matters, we would not be moving to a new basis for special concern; we would merely be extending our 'identity-based concern.' We can therefore claim that, even in these cases, identity would still be what matters.

At one point, Johnston suggests that all cases of indeterminacy can be reasonably regarded as if they involved determinate identity. But this cannot be his considered view. Johnston thinks it of great importance whether some future person would, determinately, *not* be us. If that is true, it must have some importance whether some future person would determinately *be* us. If it would matter greatly that we would be determinately dead, it must also matter that we would not determinately

still be alive.

If identity is what matters, indeterminate cases must have intermediate significance. Suppose that some future person's life would be just like ours. We can then ask whether our relation to this person would be as good as ordinary survival. If identity is what matters, and it is indeterminate whether that person would be us, our prospect must be worse than ordinary survival, but better than ordinary death.

Such a view, moreover, should not have only three values. Reconsider the *Physical Spectrum*, in which different proportions of our brain and body would be replaced, all at once, with duplicate cells. Suppose we believe that, in the last part of this range of cases, in which nearly all of our cells would be replaced, the resulting person would determinately not be us. In the first part of this range, in which there would be little replacement, the resulting person would determinately be us. On our view, in the middle of this range, there must be an indeterminate zone.³⁴

These indeterminate cases cannot all have the same intermediate significance. That would imply that, at each end of the intermediate zone, there would be a sudden change in the strength of our reason for concern. This cannot be true. The first cases of indeterminacy would be very like the last cases of determinate identity. The difference would be only that a few more of our body's cells would be replaced. And there would be as little difference between the last cases of indeterminacy and the first cases of determinate non-identity.

Within the indeterminate zone, we must apply a sliding scale. If identity is what matters, how good our prospects are, or how strong our reason is for special concern, must roughly correspond to where, in the indeterminate zone, these cases come. As we move to cases that are further from determinate identity, what matters must gradually fade away.

When he consider such cases, Johnston accepts this *gradualist* account. If it were indeterminate whether some future person would be us, our concern about that person, Johnston claims, should correspond to the strength of our 'mental and physical connections' to that person.

It may seem that, in making that claim, Johnston is abandoning his view that identity is what matters. He may seem to be conceding that, in such cases, what matters would not be identity, but these other mental and physical connections. But that is not so. If we would have more reason for concern, when these connections would be stronger, that need

not be because these connections are themselves what matter. It could be because, when these connections would be stronger, our relation to that future person would be closer to determinate identity. This gradualist account does not conflict with the view that identity is what matters. On the contrary, in cases of indeterminacy, it is the only reasonable version of that view.

This gradualist account might be expressed in a different way. Numerical identity cannot itself have degrees. A future person cannot be, to some intermediate degree, one and the same person as us. But, if we claimed that some person would be us, our claim might be said to have different degrees of truth. Consider those indeterminate cases which are close to the zone of determinate identity. In such cases, it would be *nearly* true that the resulting person would be us. At the other end of the indeterminate zone, in cases that were close to determinate non-identity, such a claim would be nearly false.

Johnston's annexation strategy could now appeal to such claims. I argued that in certain cases, though some future person would not be me, my relation to that person would contain what matters. Johnston could argue that, though it would not be determinately true that this person would be me, that would be nearly true. My relation to that person would be *close* to identity. This could be why, even if this relation would contain what matters, that does not count against the view that identity is what matters.

Is this a good reply?

11 Indeterminate Cases

Reconsider *My Division*. Johnston believes that, in this case, it is indeterminate whether the resulting people would be me. And he believes that, despite this indeterminacy, I could reasonably care about these people just as much as I would care about my own future. On the annexation strategy, we admit that

(1) my relation to each of these people would not determinately be identity,

and

(2) this relation would contain what matters,

but we also claim that

(3) this relation would be close to identity.

We then argue that, since (3) is true, (1) and (2) do not count against the view that identity is what matters.

Can Johnston defend (3)? Would my relation to each of these people be close to identity?

Such questions are easier to answer when we are considering cases in the Physical or Combined Spectra. In such a range of cases, the relevant facts or relations hold in smoothly varying degrees. Suppose we claim that, in one of these cases, the resulting person would be us. We have seen how such a claim might be nearly true. That would be so if the case in question, though in the indeterminate zone, was close to the zone of determinate identity. In *My Division*, however, there is indeterminacy for a different reason. The problem is not that, between me and some future person, the physical or mental connections hold to some reduced degree. The problem is rather that these connections take a branching form, holding between me now and each of two future people. In such a case, it is less clear how some claim about identity might be nearly true. Since the relevant facts are not here matters of degree, it is less clear what the criterion of nearness should be.

We are asking whether, in *My Division*, each of the resulting people would be close to being me. For that to be so, it must be nearly true that

(4) each of these people would be me.

Could Johnston claim that (4) is nearly true?

It would be easier to defend that claim if (4) were indeterminate. But, as Johnston would agree, (4) is false. Even false claims may be nearly true. (4), however, is self-contradictory. According to (4), two different people would each be one and the same person. And, when some claim involves a contradiction, it seems *very* false. It seems as far as it could be from being true.

There is another ground for doubting whether (4) is nearly true. If I would not be each of the resulting people, there are three other obvious possibilities. I might be either

(5) one of these people,

or

(6) the other,

or

(7) neither.

But (5) and (6) are not real possibilities. Since there are no Cartesian Egos, or other such entities, there is nothing that could make me one of these people, rather than the other. If put forward as straightforward truths, neither (5) nor (6) could be rationally believed. Nor could either be put forward as an acceptable redescription, or conceptual revision. It would be quite arbitrary to pick one of the two resulting people, and claim that he was me.

Since I cannot be each of these people, and we cannot seriously suggest that I would be one of them, we may seem entitled to conclude that, as (7) claims, I would be neither of these people. That is what several writers conclude. And, if it is determinately true that I would be neither of these people, that counts against the claim that these people would be close to being me.

My conclusion was more cautious. If (4) to (7) were the only possible descriptions, (7) might be determinately true. But there are other possible descriptions.³⁵ Partly for that reason, I agree with Johnston that there is indeterminacy here. If I were about to divide, there would be no answer to the question of what would happen to me.

Johnston and I also agree that, even without an answer, we could know what was going to happen. As he writes, there would be no 'fact of the matter' that was hidden from us. The different answers to our question do not describe different possibilities, one of which might be the truth. They are merely different descriptions of the same course of events.

Unlike Johnston, though, I make a further claim. In some indeterminate cases, such as those in the middle of the *Physical Spectrum*, there is little point in resolving the indeterminacy, by deciding to adopt one of the possible descriptions. Any such decision would be ill-grounded, and arbitrary. But, in cases of Division, things are different. If we decide to give an answer to our question, by refining our concept of a person, I believe that one description would be clearly best. Since the two resulting people cannot each be me, and it would be arbitrary to call either me, it would be best to adopt (7), and call neither of these people me.

If it would be best to say that I would be neither of these people, that description would be close to being true. And, as before, that would count against the claim that these people would be close to being me.

It may be objected that, if identity were what matters, (7) would *not* be the best description. (7) would then imply, implausibly, that Division would be not much better than ordinary death. But, in recommending (7), I cannot assume that identity is *not* what matters, since that is what my argument is trying to prove. It may therefore seem that, in defending my argument, I cannot appeal to the claim that the best description would be (7), since I would then be presupposing what I am trying to prove.

This objection is invalid. It is true that, in recommending (7), I cannot assume that identity is not what matters. Equally though, in rejecting (7), we cannot assume that identity *is* what matters. That would also beg the question since, if my argument is sound, identity is not what matters. In assessing my argument, we should not appeal to either view about identity's importance. We should ask whether, in neutral non-evaluative terms, (7) would be the best description of this case.

I believe that, in neutral terms, (7) would be best. In those cases of symmetrical division that have no rational or moral significance, we all accept claims like (7). When some amoeba splits, or some lump of clay is equally divided, we call the result two amoebae, or two lumps, neither of which is one and the same as the original.

We began by asking whether, though I could not be each of the resulting people, my relation to these people would be close to identity. I have suggested two grounds for answering No. It is, in one sense, very false that I would be each of these people, since that claim involves a contradiction. And it seems to be nearly true that I would be neither of these people.

We should next remember the purpose of our question. Johnston suggests that, even if

(1) my relation to each of these people would not determinately be identity,

and

(2) this relation would contain what matters,

we can claim that

(3) because this relation would be close to identity, (1) and (2) do not count against the view that identity is what matters.

I have described two ways in which my relation to these people does *not* seem close to identity. I could not conceivably be both of these people, and it is nearly true that I would be neither of them. Could Johnston claim that, in some other and more relevant way, these people would be close to being me? ³⁶

My relation to these people, Johnston suggests, would be a 'good surrogate' for identity. ³⁷ This relation would be close to identity in the sense of being an acceptable substitute, which would have the same importance. But *this* similarity cannot defend the view that identity is what matters. On the contrary, it counts against that view.

There is another similarity, which may seem more relevant. My relation to these people, or what I have here called *M*, combines two elements. Each of these people would be psychologically continuous with me, and this continuity would have its normal cause: the continued existence of enough of my brain. This is the relation, Johnston claims, which in ordinary cases constitutes identity, or the continued existence of the same person. In many indeterminate cases, this relation would hold only in a weaker form, with a different cause, or with some elements missing. In *My Division*, in contrast, nothing would be missing. Though I cannot be identical to each of the resulting people, that relation fails to be identity only for extrinsic reasons. But for the existence of the other person, each of the resulting people would *be* me.

I concede that, in *this* sense, each of these people would be very close to being me. But, as before, this fact cannot defend the view that identity is what matters.

Return to my imagined community who believe that it matters greatly that each of them should have a twin. Suppose that one of these people challenges that belief. What really matters, this person claims, is not having a twin, but having at least one sibling. To support this view, this person argues that, if these people each had two siblings, who were their triplets, that would be just as good. Suppose that the other people accept this claim. Someone then objects that, if they had two siblings, these siblings would be *close* to being their twins. Each after all, would *be* their twin if the other did not exist. That reply would have no force. Having a sibling is, indeed, close to having a twin. In this community, the first relation always constitutes the second. But this cannot defend the view that what matters is having a twin *rather than* having at least one sibling.

Similar remarks apply to our discussion. What matters, I argue, is not personal identity, or the fact that some future person will be us. It is rather that at least one future person will be M-related to us. To support this view, I appeal to a case in which, because two future people would be M-related to me, it is not true that either person would be me. Suppose we agree that, in this case, my relation to these people would contain what matters. We might object that this relation would be close to identity, since, if it held in a one-one form, it would constitute identity. But this claim is irrelevant. However close these relations are, this cannot defend the view that what matters is identity *rather than* this other relation.

12 Does M have only derivative importance?

Johnston might now appeal to another of his claims. He suggests that even if, in this example, M would contain what matters, M would not be important in its own right. M's importance would be parasitic, since it would be derived from identity's importance. If Johnston could defend this claim, that would undermine my argument.

Can Johnston defend this claim? When he explains why M would matter, Johnston writes:

the important core of what constitutes identity is still discernible in the fission case, i.e. the persistence of enough of the brain to be capable of continuing the mental life of the original subject. This important core also holds twice over in the fission case. So were we ever to face fission it would be reasonable to care about our fission products as we would care about a future self.³⁸

This passage takes us back to our earlier and central disagreement. Johnston talks of 'the important core of what constitutes identity'. On my view, it is this core itself which is important. What prudentially matters is not identity, but some of the relations which, in ordinary cases, constitute identity. On Johnston's view, these other relations are not important in themselves. They are important only because, in ordinary cases, they constitute identity.

This disagreement raises general questions. My book claimed that, when one fact consists in certain others, it must be these other facts which matter. As Johnston shows, that was a mistake. I should have appealed to the different claim which I have here called Realism about Importance.

There is another way to compare Johnston's view and mine. In ordinary cases, in which M constitutes identity, M and identity together have certain kinds of importance. Because these relations here coincide, it is hard to judge their relative importance. We can therefore turn to cases where M and identity diverge. We should ask whether, in such cases, M by itself would have the importance that, in ordinary cases, M and identity together have.

We have been considering one such case. In *My Division*, though I would be M-related to each of the resulting people, this relation would not constitute identity. We might decide that, since these people would not be me, my relation to them would not contain what matters. If that were so, this example would support Johnston's view. If M mattered only *when* it constituted identity, that would suggest that M matters only *because* it constitutes identity. M's importance would then be derivative, and identity would be what really matters. As Johnston seems to agree, however, my relation to these people *would* contain what matters. This relation would have the importance that, in ordinary cases, M and identity together have. This fact supports my view. It suggests that M is what matters.

Since this is only one example, we should ask whether there are other cases in which M and identity diverge. Suppose that, in all such cases, we reach similar conclusions. Suppose we decide that

(8) whenever M and identity diverged, M would be what mattered.

This would give us further reason to accept my view.

We are now discussing one possible reply. Johnston suggests that, in these cases, M would not matter in its own right. Even if M by itself would contain what matters, M's importance would be derivative, or parasitic. M would contain what matters, in these cases, only because, in ordinary cases, it constitutes identity.

Such a view is coherent. But it seems implausible. If M is important only because it constitutes identity, why would it have the same importance even when it did not constitute identity? If identity is what matters, why would it make no difference whether identity was there?

It may help to consider other, similar disagreements. There is a strong objection, most of us believe, to the killing of innocent people. But this belief, some people argue, needs to be revised. On their view, the real objection is not to killing, but to a wider class of acts: those that

intentionally, in certain ways, cause people's deaths.

We have discussed one case in which these views seem to disagree. Imagine that White poisons you, and that, because Grey knows that you are about to die from White's poison, Grey saves someone else's life in a way that kills you. Though White intentionally causes your death, White does not kill you. It is Grey who, permissibly, does that. Suppose we believe that, though White does not kill you, this fact is morally irrelevant. We believe that what White does is as bad as killing. This conclusion would support the second view described above. It suggests that the true objection is to acts that intentionally, in certain ways, cause deaths. Suppose next that, in all other cases of this kind, we have similar beliefs. We think it just as bad intentionally to cause such deaths, even when these acts do not constitute killings. We would then have strong reason to accept the second view.

Holders of the first view have one possible reply. They might say: 'The true objection *is* to killing. Such intentional causings of death may be as bad as killing; but that badness is derivative. Such acts are bad only because, in most cases, such intentional causings of death constitute killings.'

This reply, though coherent, is not plausible. If such acts are bad only because, in most cases, they constitute killings, why are they just as bad even when they do not constitute killings? If the real objection is to killing, why does it make no moral difference whether these acts kill?

Consider next another pair of views about the morality of killing. According to some people, while there may be some objection to killing other animals, there is a much stronger objection to the killing of human beings. According to some other people, the stronger objection is not to killing all human beings; it is to killing only those human beings who have developed to the stage at which they have brains, and nervous systems.

To decide between these views, we might ask what we believe about the killing of a fertilized human ovum, or an early foetus. We might believe that there is no strong objection to the killing of such undeveloped human beings. We would then have reason to accept the second view. But suppose we give a different answer. Suppose we are convinced that the killing of an early foetus is just as bad as the killing of a developed human being. We then have reason to believe that, as the first view claims, the true objection is to killing any human being.

As before, holders of the second view might say: 'The true objection *is* to

killing developed human beings. Though it is just as bad to kill undeveloped human beings, this badness is derivative. Such acts are bad only because, in most cases, the killing of a human being is the killing of a developed human being.'

This reply, as before, is not plausible. If the true objection is to killing developed human beings, why is it just as bad to kill undeveloped human beings? If this distinction is morally important, why does it make no moral difference?

Similar remarks apply to Johnston's view. We are now assuming that, even when M did not constitute identity, M would contain what prudentially matters. M by itself would have the same importance that, when it constitutes identity, M and identity together have. This supports the view that, of the two, M is what matters. Johnston replies that, in these special cases, M's importance would be derivative. M would contain what matters, even when it did not constitute identity, only because in ordinary cases it does constitute identity.

This view, I have said, faces similar objections. If M is important only because it constitutes identity, why would it have the same importance even when this was *not* true? If identity is what matters, how can M all by itself matter as much as identity-and-M? Why would it not matter that identity was not preserved?

Johnston's view does, in a way, answer these questions. On this view, though identity is what really matters, M has derivative importance because, in ordinary cases, it constitutes identity. Though this importance is derivative, it is equal to identity's importance, and M would keep this importance even in cases where identity was not preserved. That might be said to explain how, in such cases, M by itself would matter as much as identity and M.

These claims do not, however, answer the objections. Our question now becomes: why should we believe these claims? If M by itself would have the same importance, why should we believe that M's importance is derivative? If M would matter even in cases in which it did not constitute identity, why believe that M matters only because, in *other* cases, it does constitute identity?

Johnston's answer might be the following. Identity is what matters, he believes, because it provides the actual basis for our future-directed concern. We care about our future because it will be *our* future. It may be true that, in some imagined cases in which identity and M diverged, M

would have all of identity's importance. In such cases, it might be M that gave us reason for special concern. But this would be only a *local* change in the basis for our concern: one that would be restricted to these special cases. In ordinary life, Johnston claims, there would be 'no effect upon the reasonableness of organizing our concern in terms of identity.' M would *not*, he concludes,

be the relation around which our self-referential concerns are almost always organized. Identity would keep this privileged role. Within locally modified concern the importance of M. . . in a few cases would be parasitic on the importance of identity.

This reasoning is, I think, unsound. Johnston claims that

(9) in ordinary cases, in which M and identity coincide, it is reasonable to organize our concerns in terms of identity.

That does not show that

(10) M's importance is parasitic on identity's importance.

Suppose we would be right to conclude that, in all cases, M is what really matters. It might still be reasonable, in ordinary cases and our actual lives, to organize our concerns around identity. This would not help to show that M's importance is derived from identity's importance. On the contrary, on these assumptions, identity's importance would be derived from M's importance. If it were reasonable to organize our concerns around identity, that would only be because, in ordinary life, identity always coincides with M. Given that coincidence, even if M were what really matters, we might have no reason to revise our identity-based concern. Such concern would never in fact go astray.

I believe that, if M is what matters, our identity-based concern does in some ways go astray. I would therefore question (9). But this does not affect the point that I have just made. Even if (9) were true, that would not justify (10). It is often reasonable to organize our concerns, not in terms of what really matters, but in terms of some other simpler correlate of what matters. (9) could be true even if identity were never, in itself, what mattered.

There may be some other reason to accept (10). But Johnston's paper does not, I believe, give us such a reason. And we may, I have claimed, have a reason to reject (10). We may decide that

(8) whenever M and identity diverged, M would be what mattered.

If that is so, we could reasonably conclude that M is always what matters. It would then be identity's importance that would be parasitic. The truth would not be that M matters because it constitutes identity. Rather, identity matters because it consists in non-branching M.

13 The Divergence Argument

Return now to my Divergence Argument. I claimed:

(1) In ordinary cases, if I would be M-related to some future person, this person would be me. In these cases, M and identity coincide.

(2) In *My Division*, M and identity would diverge. Though I would be M-related to each of the two resulting people, it would not be true that each of them is me.

(3) My relation to each of these people would contain what mattered.

(4) If there were other kinds of case in which M and identity diverged, these would also be cases in which it would be M that contained what mattered.

Therefore

(5) Whenever M and identity diverged, it would be M that contained what mattered.

Therefore

(6) Even when these relations coincide, it is M that contains what matters. Identity is never what matters.

(1) is uncontroversial. (2), as we have just seen, is hard to deny. *My Division* Argument shows, I believe, that we should accept (3). (1) to (4) together entail (5). And I have defended the inference from (5) to (6). So we have one remaining question. Should we accept (4)? Are there other kinds of case in which M and identity diverge? If so, would these also be cases in which it would be M that mattered?

There are two ways in which M and identity might diverge. I might either

(a) be M-related to some future person who would not be me,

or

(b) fail to be M-related to myself in the future.

(a) would be true in *My Division* because I would be M-related to *more than one* future person. That is one way in which, according to nearly all criteria of identity, M and identity could diverge. We can call these *branching* cases. Whether there could be other cases depends on which criterion of identity we should accept. Since that is controversial, we can review the main alternatives.

Suppose, first, that we accept some form of the Narrow or Brain-Based Psychological Criterion. On this view, a future person would be me if and only if we would be uniquely M-related. Identity consists in non-branching M. If that is so, it can only be in branching cases that M and identity diverge. This, by default, establishes premise (4). If we accept this Narrow Criterion, we should therefore now accept all of my argument's premises. In branching cases, M would contain what mattered. On our view, there could not be other cases in which M and identity diverged. So, in all cases of divergence, it would be M that contained what mattered. We could thus infer that, even in ordinary cases, in which M and identity coincide, it is M and not identity that contains what matters.

Suppose next that we reject the Narrow Criterion. In discussing other criteria, we can remember that M has two main elements. I shall be M-related to some future person if we are psychologically continuous, and this continuity will have its normal cause: the continued existence of enough of my brain.

On the Wide Psychological Criterion, our continued existence requires only the first of these two elements. A future person will be me if and only if we are uniquely psychologically continuous. As before, on this view, there are branching cases of type (a) in which M and identity diverge. But there is now a second possibility. There might be some future person who will be uniquely psychologically continuous with me, but without having any of my brain. These would be cases of type (b). In such cases, on the Wide Criterion, I would fail to be M-related to myself in the future. One such case is *Teletransportation*, in which my body would be destroyed and replicated.

If we accept this Wide Criterion, what should we believe to be what

matters? We should again conclude that, in branching cases, it would be M that contained what mattered. But should I agree to be teletransported? On our view, though I would not be M-related to the person who woke up in the cubicle on Mars, this person would be me. Would my relation to myself on Mars contain what mattered?

If we answered No, we would have further reason to believe that identity is not what matters. But, of those who accept this criterion, many would answer Yes. If we give this answer, we might conclude that, in this example, identity would be what mattered. I should agree to be teletransported, we might claim, because this would get *me* to Mars.

If we draw this conclusion, though, we would have an oddly hybrid view. On this view, there would be what matters in both

(i) my relation, in *My Division*, to each of the two resulting people,

and

(ii) my relation to the person who woke up on Mars.

(i) would contain what matters though this relation is *not* identity, but (ii) would contain what matters because this relation *is* identity. Though such a view might be true, it would be puzzling. Why would identity's importance vary in this way?

There is another, simpler view. If there would be what matters in both (i) and (ii), what matters, we can assume, would be the same in both cases, and would be contained in what these relations have in common. This common element is not identity, since I could not be each of the two resulting people. Nor is it M, since I would not be M-related to the person who woke up on Mars. What these relations have in common is one element in M: psychological continuity. Let us call this *relation C*. I would be C-related to all three people. On the assumptions we are now making, we should conclude that what mattered would be, or be contained in, C.

There is another argument for this conclusion. Suppose that, in *Double Teletransportation*, after my body is destroyed, it would be replicated twice. On the Wide Criterion, this would be merely another version of *My Division*. We can therefore reapply my Division Argument. My relation to either Replica would be intrinsically the same as my relation, in *Single Teletransportation*, to myself on Mars. On our assumptions, though this relation would not be identity, it must contain what matters. So we

should think the same about Single *Teletransportation*. I should agree to be teletransported, not because my Replica would be me, but because we would be psychologically continuous.

Remember next that, on the Wide Criterion, our identity consists in non-branching C. So it is only in branching cases that C and identity could diverge. We could therefore claim that, whenever C and identity diverged, it would be C and not identity that contained what mattered. And we could infer that, even in ordinary cases, that is true.

Suppose, next, that we accept the Brain Criterion. This appeals to the other element in M: the continued existence of enough of the same brain. We can call this *relation B*. On the Brain Criterion, a future person will be me if and only if we are uniquely B-related.

Similar remarks apply. On the Brain Criterion, there are two kinds of case in which M and identity could diverge. As before, there are branching cases, like *My Division*. There could also be cases of type (b), in which I would fail to be M-related to myself in the future. One example is Williams's imagined *Brain Zap*, in which the remodelling of my brain would destroy all psychological connections, and would give the resulting person wholly new psychological features. On the Brain Criterion, this person would be me.

If we accept the Brain Criterion, what should we believe to be what matters? We have the same reasons to conclude that, in *My Division*, it would be M and not identity that contained what mattered. But what should we think of Williams's example? A brain zap might destroy what, in the desirability sense, matters. But that is not our question. Would a brain zap destroy what, in the prudential sense, matters? If I were about to have this operation, would I have reason for special concern about my future?

Some of us would answer No. If we accept this answer, we would have further reason to believe that identity is not what matters. But suppose we answer Yes. Suppose we believe that, even if I were about to lose all of my psychological features, I would still have reason for such concern. Williams's discussion can make that seem compelling. If we accept this second answer, we might conclude that identity would here be mattered. I would have reason for special concern, we might claim, because this would be *my* future.

This conclusion would, though, be at odds with our conclusion about *My Division*. According to these two conclusions, there would be what

matters in both

(i) my relation, in *My Division*, to each of the two resulting people,

and

(iii) my relation to my brain-zapped future self.

(i) would contain what matters, though this relation is not identity, but (iii) would contain what matters because this relation is identity. That would be puzzling. We should ask instead what these relations have in common. That common element, we can assume, would contain what matters. That element is not identity, or M, but B: having enough of the same brain. Given our beliefs about these two cases, we should conclude that, in both cases, what mattered would be, or be contained in, B.

There is another argument for this conclusion. On the Brain Criterion, I could survive if enough of my brain were successfully transplanted. And I could survive if my brain were zapped. So there could be a brain-zapped version of *My Division*. If half my brain were destroyed, and the other half were zapped and transplanted, the resulting person would be me. On our present assumptions, my relation to that person would contain what mattered. If both halves were zapped and transplanted, neither resulting person would be me. But my relation to each person would be, in itself, the same; so, on our assumptions, it would still contain what mattered. We should therefore think the same about Williams's original example. I should be specially concerned my brain-zapped future, not because it would be my future, but because I would still have enough of the same brain.

Note next that, on the Brain Criterion, it is only in branching cases that B and identity could diverge. We could therefore claim that, whenever these relations diverged, it would be B that contained what mattered. And we could infer that, even in ordinary cases, that is true.

Suppose finally that, like Williams, we accept the Bodily Criterion. Similar claims would apply. Since these claims are similar, I give them only in an Appendix.

In my discussion of these four criteria, there has been a recurring pattern. This suggests another version of the Divergence Argument. We can claim:

(1) On any plausible criterion, our identity consists in the non-branching form of some *grounding* relation, *G*. Some future person will be me if and only if we are uniquely *G*-related.

(2) In branching cases, in which *G* and identity diverge, what mattered could not be identity, though it might be *G*.

(3) Since identity consists in non-branching *G*, these are the only cases of divergence.

Therefore

(4) Whenever *G* and identity diverged, what mattered could not be identity, though it might be *G*.

Therefore

(5) Even in ordinary cases, in which *G* and identity coincide, what matters cannot be identity, though it might be *G*.

I have defended (1) and (2), and (3) is uncontroversial. These claims entail (4).³⁹

This conclusion, it may be said, would not strongly contradict the view that identity is what matters. What matters would, after all, be our criterion's grounding relation: the relation which, except in branching cases, constitutes identity. And the exceptional, divergent cases are all purely imaginary. It is hardly surprising that, in holding our ordinary view about identity's importance, we ignore such cases.

A similar objection might be urged against my other argument for the unimportance of identity: what I have called here Realism about Importance. According to this argument, what matters cannot be identity itself, though it might be one or more of the facts or relations in which identity consists. Here too, it may be said, this claim hardly contradicts the ordinary view. Suppose that, because we are realists about importance, we conclude that that what matters is not personal identity but the facts in which such identity consists. In reaching this conclusion, we need not appeal to any claimed divergence between identity and what matters. Even if there could not be branching cases, and *G* and identity always coincided, we could still argue that it could not be identity that really mattered. But, if we concluded that it was *G* that mattered, that might be said hardly to conflict with the ordinary view. In both cases, we may misidentify the level of the fact that matters, or which of these facts have only derivative importance. But it might be

true that, in all cases where there is what matters, there is identity, and vice versa. And this coincidence might be claimed to be all that common sense maintains.

I believe, however, that on no criterion of identity is G the whole of what matters. And, on some criteria of identity, G does not even contain what matters. Partly for this reason, even in ordinary cases, what matters and identity do not wholly coincide.

It is on the Bodily Criterion that identity has least connection with what matters. On this criterion, G is W: or having enough of the same body. M, I believe, contains what matters. To ask whether that is so, we must consider cases in which M and W diverge.

Return to the case discussed above, in which my head would be destroyed, and your head would be grafted onto the rest of my body. On the Bodily Criterion, it would be I who woke up. If that is so, I argued, identity would not here be what matters. If there will be someone alive tomorrow who will have your head, your relation to this person contains what matters. It is irrelevant whether this person would be you. And if my head will be destroyed, it would not give me what matters that there will be someone alive tomorrow with the rest of my body. It is irrelevant whether this person would be me.

Suppose that we accept these claims. If we accept the Bodily Criterion, we could then argue:

- (1) In a case like *Head Transfer*, M would diverge from both identity and W.
- (2) In such cases, what mattered would not be identity, or W. It would be M that was or that contained what mattered.
- (3) If there were other cases in which M diverged from identity and W, these would also be cases in which it would be M that contained what mattered.

Therefore

- (4) Even in ordinary cases, in which M coincides with identity and W, it is M that is, or that contains, what matters.

On this conclusion, what matters is neither identity itself, nor the grounding relation, W, which in ordinary cases constitutes identity. The most that could be claimed is that, in ordinary cases, identity and

what matters coincide, or hold together.

On the objection mentioned earlier, that would be enough to vindicate the common sense view. When common sense believes that identity is what matters, it may only believe that, in all cases in which there is what matters, there is identity, and vice versa.

Common sense, however, believes more than that. There may be several other facts which coincide with what matters. If that is true of identity, it is true of such facts as having the same fingerprints, or the same belly-button. It might be true that, in ordinary cases, if some future person will have our finger-prints, our relation to that person will contain what matters. But this coincidence, though not a *mere* coincidence, could not justify the view that having the same fingerprints is what matters. And, on the Bodily Criterion, there is as little connection between identity and what matters. Neither identity itself, nor the facts in which it consists, are any part of what matters.

If, as I have argued, personal identity isn't what matters, we must ask what does matter. That is too large a question to start discussing here. There are several possible answers. And, depending on our answer, there are several further implications. Thus there are several moral questions which I shall not even mention. I shall end with another remark about our concern for our own future.

This 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 this child's 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.

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 this belief may underlie our anticipatory concern about our own future.

There are, I suspect, several causes of that illusory belief. I have discussed one such cause: 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.

APPENDIX

On the Bodily Criterion, a future person will be me if and only if this person will have my body. If we accept this view, we shall reject one premise of my original Division Argument. We shall deny that, in the *Single Case*, the person with half my brain would be me. This may lead us to deny that my relation to this person would contain what mattered.

The Bodily Criterion, as we have seen, must be revised. We can imagine how, with advanced prosthetic surgery, I might survive with only half my body. So our criterion must appeal to what we can call *relation W*: having enough of the same body. And, in defining *W*, we must allow that half could be enough. On our revised criterion, a future person will be me if and only if we are uniquely *W*-related. My argument could now appeal to a case of Bodily Division. If there was only one future person whose body consisted in half my present body, that person would be me. But, if there were two such future people, they could not each be me.

On the Bodily Criterion, there are two kinds of case in which *M* and identity could diverge. There are branching cases, like the Bodily

Division I have just described. But there are also simpler cases. One example is Head Transfer, in which my head would be destroyed, and your head would be grafted onto the rest of my body. On the Bodily Criterion, it would be I who woke up. If that is so, you would be M-related to a future person who would not be you, and I would fail to be M-related to myself in the future.

If we accept the Bodily Criterion, what should we believe to be what matters? We have the same reasons to conclude that, in a case of Bodily Division, my relation to each of the resulting people would contain what mattered. But how should we regard the replacement of my head by yours? If someone alive tomorrow would have your head and the rest of my body, would either you, or I, be related to this person in the way that mattered?

We might believe that, as I have argued, your relation to that person would contain what mattered. This would give us further reason to conclude that identity is not what matters. But suppose we accept the other answer. We believe that, since the person with your head tomorrow would be *me*, it is my relation to this person that would contain what mattered. We would then be claiming that identity would here be what mattered. As before, this would be puzzling. On our view, there would be what matters in both

(i) my relation, in my Bodily Division, to each of the two resulting people,

and

(iv) my relation to myself after your head has replaced mine.

(i) would contain what matters, though this relation is not identity, but (iv) would contain what matters because this relation is identity. We could more reasonably assume that what matters is something that is present in both (i) and (iv). This common element is not identity, or M, but W: having enough of the same body. On these assumptions, we should conclude that, in both cases, what mattered would be, or be contained in, W.

Once again, there is another argument for that conclusion. On the Bodily Criterion, I could survive if your head replaced mine. And I could survive if half my body were destroyed but the rest remained the body of a living person. So there could be a head-receiving version of my Bodily Division. On our view, if my head and half my body were destroyed,

but the rest of my body received your head, the resulting person would be me. My relation to this person, we are now assuming, would contain what mattered. But my relation to this person would be in itself the same if the other half of my body were not destroyed, but received the head of some third person. On our assumptions, this relation would still contain what mattered, even though it would not, here, be identity. We should therefore think the same about the original example, in which your head replaces mine. I should be specially concerned about the resulting person, not because this person would be me, but because he would have enough of my body.

As before, on the Bodily Criterion, it is only in branching cases that W and identity could diverge. On these assumptions, we could therefore claim that, whenever W and identity diverged, it would be W that contained what mattered. And we could infer that, even in ordinary cases, identity is not what matters.

¹ My book was published by the Oxford University Press in 1984, but was reprinted in 1987 with minor revisions. Since this paper is based on rough drafts written some years ago, it does not attempt to discuss any recent published work. The first section of this paper, and some later passages, are taken from my article 'The Unimportance of Identity' which appeared in a collection of mostly non-philosophical papers, *Identity*, edited by Henry Harris (Oxford University Press, 1995). Mark Johnston's objections to my arguments are in his article 'Human Concerns without Superlative Selves', in *Reading Parfit*, edited by Jonathan Dancy (Blackwell, 1997). I have published two other long papers on personal identity: 'Experiences, Subjects, and Conceptual Schemes' in *Philosophical Topics*, Vol 26, No 1 and 2, Spring and Fall 1999, and 'Persons, Bodies, and Human Beings', in *Contemporary Debates in Metaphysics*, edited by John Hawthorne, Dean Zimmerman, and Theodore Sider (Blackwell, 2008).

² Vasubandhu, quoted in Theodore Stcherbatsky, 'The Soul Theory of the Buddhists', *Bulletin de l'Academie des Sciences de Russie*, 1919, p 485.

³ The Visuddhimagga, quoted in Steven Collins, *Selfless Persons* (Cambridge University Press, 1982).

⁴ Even in cases involving nations or machines, there are some deep philosophical problems raised by indeterminacy, which I shall here ignore. For my purpose, it is enough to distinguish between cases of indeterminacy that do or don't seem puzzling to most non-philosophers.

⁵ By 'what matters' I didn't mean 'what actually matters to us', i.e. what we take to matter. It cannot be argued that, in this sense, identity is not what matters. I meant what we have *reason* to care about or should believe to matter.

⁶ In *Climbing the Mountain* (Oxford University Press, in press).

⁷ Bernard Williams, 'The Self and the Future', *Philosophical Review* 1970, reprinted in Williams's *Problems of the Self* (Cambridge University Press, 1973).

⁸ And, if I disliked my character, and other features, I might agree. I might welcome such a chance to throw myself away and start again.

⁹ Robert Adams 'Existence, Self-interest and the Problem of Evil' *Nous* 13, 1979.

¹⁰ Peter Unger, *Identity, Consciousness, and Value* (Oxford University Press, 1990).

¹¹ John Perry, 'The Importance of Being Self-Identical', in ed. Amelie Rorty *The Identities of Persons* (University of California Press, 1976).

¹² See, for example, Jennifer Whiting, 'Friends and Future Selves', *The Philosophical Review*, 1986

¹³ *Reading Parfit*, 159.

¹⁴ Susan Wolf, for example, suggests such a view in her 'Self-Interest and Interest in Selves', *Ethics* July 1986. I offer a brief response in my *Comments* in the same issue.

¹⁵ Robert Adams 'Should Ethics Be More Impersonal?' *The Philosophical Review* 1989, reprinted in *Reading Parfit*, 264.

¹⁶ In my 'Comments' in *Ethics* July, 1986.

¹⁷ To give some examples, the claim that identity is not what matters is rejected, in *Reading Parfit, op.cit.* by Robert Adams, Mark Johnston, David Brink, John McDowell, Judith Thomson, and Simon Blackburn. Only Shoemaker agrees with me.

¹⁸ In his 'Human Concerns without Superlative Selves', in *Reading Parfit, op.cit.* Future references are to this article.

¹⁹ As it happens, Johnston accepts my view about this question. But that is irrelevant. We are now asking whether, as Johnston claims, we should accept this kind of argument.

²⁰ 168.

²¹ See, for example, Michael Ayers, Locke, Vol II, Paul Snowdon, *Persons, Animals, and Ourselves* in *The Person and the Human Mind*, edited by Christopher Gill (Oxford University Press, 1990).

²² See, for example, Thomas Nagel, *The View from Nowhere* (Oxford University Press, 1986) 40-45, John Mackie, *Problems from Locke* (Oxford University Press, 1976) Chapter 6, and Jefferson McMahan *Killing at the Margins of Life* (Oxford University Press, 2002).

²³ In 'Persons, Bodies, and Human Beings', in *Contemporary Debates in Metaphysics*, edited by John Hawthorne, Dean Zimmerman, and Theodore Sider (Blackwell, 2008).

²⁴ If we think that, in all these cases, this person would be me, we could turn instead to the cases in the middle of the *Combined Spectrum*.

²⁵ 168-9.

²⁶ Ernest Sosa 'Surviving Matters' *Nous* 24 1990.

²⁷ Susan Wolf 'Self-Interest and Interest in Selves', *Ethics* 96 1986.

²⁸ 169.

²⁹ 171.

³⁰ My book defined R as psychological continuity *and/or* connectedness, with the right kind of cause. That was inaccurate, since this definition undermines my claim that personal identity consists in non-branching R. Since my definition contains the phrase 'and/or', that implies that our identity might consist in non-branching psychological connectedness. But, as I point out, a criterion of identity must appeal to a relation which, unlike connectedness, is transitive. What I mean by R here is psychological continuity *and* connectedness, and I shall also assume, for the rest of this section, that this relation has its normal cause, the continued existence of enough of the same brain.

³¹ 171.

³² 160.

³³ 171.

³⁴ It is not clear that there would be a first case of indeterminacy. Such cases raise deep problems, which may not have been solved. But we can ignore these problems here.

³⁵ Some of which I discuss in *Reasons and Persons*, Section 89.

³⁶ These remarks are not conclusive. Though contradictions are, in one sense, very false, there may be some other sense in which they are close to being true. Consider the claim that 1,000,001 equals 1,000,002. Though that involves a contradiction, it is nearly true.

³⁷ 166.

³⁸ 170.

³⁹ The inference from (4) to (5), though plausible, needs one more premise. (2) leaves it open whether, in branching cases, what mattered would be G. Suppose that what mattered would not be G, but some other relation, which we can call X. This other relation might be some element of G, or it might merely, in these cases, coincide with G. There might then be other cases where X and identity diverged; and, in those cases, it might not be X that mattered. To complete this argument, we would then have show that there were no such cases. We would have to claim

(6) There is some relation X such that, whenever X and identity diverged, it would be X that mattered.

Only then could we infer that, even when X and identity coincide, it is X and not identity that matters. We may not need this extra premise. Suppose we decide that, in branching cases, G *would* be what mattered. G would then be the X that satisfied (6), and this argument would be complete. We could infer that, in all cases, G is what matters.