

Why we are not what we believe

Jeff Speaks
PHIL 30304

November 6, 2018

1	Arguments against Cartesianism (§82)	1
1.1	The lack of evidence	1
1.2	The featureless Cartesian view	2
2	Arguments against the psychological criterion	2
2.1	Circularity (§80)	2
2.2	Experiences and subjects (§81)	3
3	Spectrum arguments	4
3.1	Willams and torture (§83)	4
3.2	The psychological spectrum (§84)	5
3.3	The physical spectrum (§85)	5
3.4	The combined spectrum (§86)	6

In this chapter, Parfit discusses a number of different theories of personal identity. It will be useful to discuss these views in a slightly different order than he does.

1 Arguments against Cartesianism (§82)

While most of the chapter is devoted to discussion of reductionist views, he briefly gives some reason for doubting that non-reductionist views like Descartes' are true.

1.1 *The lack of evidence*

Parfit thinks that Descartes' theory of what we are is intelligible, but that we lack any evidence for it. He discusses two different kinds of possible evidence:

- (1) Evidence from reincarnated subjects.
- (2) Cases in which the connection between psychology and the brain is, in various ways, irregular.

But, Parfit concludes,

‘We do not in fact have the kind of evidence described above. Even if we can understand the concept of a Cartesian Pure Ego, or spiritual substance, we do not have evidence to believe that such entities exist. Nor do we have evidence to believe that a person is any other kind of separately existing entity.’

How strong is this argument? How would Descartes reply?

1.2 *The featureless Cartesian view*

Parfit gives another argument against the Cartesian view, which is derived from Locke and Kant. This second argument can be presented as a dilemma. Should the Cartesian say that there is a necessary connection between immaterial souls and happenings in the material world, or not?

If we say yes, this would seem to undercut the main arguments for the dualist view.

So suppose we say no. This, Parfit thinks, opens up the following possibility:

‘the Cartesian Ego that I am might suddenly cease to exist and be replaced by another Ego. This new Ego might ‘inherit’ all of my psychological characteristics, as in a relay race. On this Featureless Cartesian View, while you are reading this page of text, you might suddenly cease to exist, and your body be taken over by some new person who is merely exactly like you. If this happened, no one would notice any difference. There would never be any evidence, public or private, showing whether or not this happens, and, if so, how often. We therefore cannot even claim that it is unlikely to happen. And there are other possibilities. On this view, history might have gone just as it did, except that I was Napoleon and he was me. This is not the claim that Derek Parfit might have been Napoleon. The claim is rather that I am one Cartesian Ego, and that Napoleon was another, and that these two Egos might have ‘Occupied’ each other’s places.’

If this possibility is accepted, Parfit thinks, the dualist view becomes ‘unintelligible.’ The argument he has in mind might be presented like this:

1. If dualism is true, then ‘swapping’ examples like the one described are possible.
 2. These examples are not possible.
-
- C. Dualism is false.

How good is this argument?

2 Arguments against the psychological criterion

Parfit's discussion of reductionist views can be divided into two categories. First, he discusses and responds to various arguments against the psychological criterion. Second, he discusses what he calls 'spectrum' arguments. Let's look at these in turn.

2.1 *Circularity* (§80)

Parfit puts the first objection, which is derived from Butler, like this:

'It is part of our concept of memory that we can remember only our own experiences. The continuity of memory therefore presupposes personal identity. The same is therefore true of your Relation R. You claim that personal identity just consists in the holding of Relation R. This must be false if Relation R itself presupposes personal identity.'

His reply: defining the notion of a quasi-memory:

'I have an accurate quasi-memory of a past experience if (1) I seem to remember having an experience, (2) someone did have this experience, and (3) my apparent memory is causally dependent, in the right kind of way, on that past experience.'

There is no contradiction in supposing that I quasi-remember someone else's experiences, so quasi-memory does not presuppose facts about personal identity.

2.2 *Experiences and subjects* (§81)

A second objection is that the very existence of experiences presupposes the existence of a separately existing subject, who is the subject of those experiences. Parfit puts the second objection like this:

'In memory we are directly aware of our own identity through time, and aware that this is a separate, further fact, which cannot just consist in physical and psychological continuity. We are aware that each of us is a persisting subject of experiences, a separately existing entity that is not our brain or body. And we are aware that our own continued existence is, simply, the continued existence of this subject of experiences.'

Parfit replies to this directly:

‘I do not believe that I am directly aware that I am such an entity. And I assume that I am not unusual. I believe that no one is directly aware of such a fact.’

His main argument for this concerns the Branch Line case. Consider your Replica on Mars. S/he will seem to remember various experiences of yours, and may even claim to be aware of a separately existing entity which has existed for a long time. But these are mistakes. So your Replica is not immediately aware of a subject of experiences which has persisted throughout a series of quasi-memories. So our experiences and conscious life give us no particular reason to think that there is such a persisting subject of experience.

3 Spectrum arguments

3.1 *Williams and torture (§83)*

Parfit later turns to what he takes to be a more serious objection to the psychological criterion, which is based on an example due to Bernard Williams. He gives the following version of the case:

‘I am the prisoner of some callous neurosurgeon, who intends to disrupt my psychological continuity by tampering with my brain. I shall be conscious while he operates, and in pain. I therefore dread what is coming. The surgeon tells me that, while I am in pain, he will do several things. He will first activate some neurodes that will give me amnesia. I shall suddenly lose all of my memories of my life up to the start of my pain. Does this give me less reason to dread what is coming? Can I assume that, when the surgeon flips this switch, my pain will suddenly cease? Surely not. The pain might so occupy my mind that I would even fail to notice the loss of all these memories.’

The surgeon next tells me that, while I am still in pain, he will later flip another switch, that will cause me to believe that I am Napoleon, and will give me apparent memories of Napoleon’s life. Can I assume that this will cause my pain to cease? The natural answer is again No. To support this answer, we can again suppose that my pain will prevent me from noticing anything. I shall not notice my coming to believe that I am Napoleon, and my acquiring a whole new set of apparent memories. When the surgeon flips this second switch, there will be no change at all in what I am conscious of. The changes will be purely dispositional. It will only become true that, if my pain ceased, so that I could think, I would answer the question ‘Who are you?’ with the name ‘Napoleon’. Similarly, if my pain ceased, I would then start to have delusory apparent memories, such as those of reviewing the Imperial Guard, or of weeping with frustration at the catastrophe of 1812. If it is only such changes in

my dispositions that would be brought about by the flipping of the second switch, I would have no reason to expect this to cause my pain to cease.

The surgeon then tells me that, during my ordeal, he will later flip a third switch, that will change my character so that it becomes just like Napoleon's. Once again, I seem to have no reason to expect the flipping of this switch to end my pain. It might at most bring some relief, if Napoleon's character, compared with mine, involved more fortitude.'

What will the proponent of the psychological criterion say about this case? Why does this seem implausible to Parfit?

3.2 *The psychological spectrum (§84)*

Parfit then gives a more general version of Williams' argument, which he calls the psychological spectrum:

'In the case at the far end, the surgeon would cause very many switches to be simultaneously flipped. This would cause there to be no psychological connections between me and the resulting person. This person would be wholly like Napoleon.

In the cases at the near end, the surgeon would cause to be flipped only a few switches. If he flipped only the first switch, this would merely cause me to lose a few memories, and to have a few apparent memories that fit the life of Napoleon. If he flipped the first two switches, I would merely lose a few more memories, and have a few more of these new apparent memories. Only if he flipped all of the switches would I lose all my memories, and have a complete set of Napoleonic delusions.'

The proponent of the psychological criterion seems forced to say that there is some place in the spectrum at which one would cease to exist. But, Parfit says, this is very hard to believe.

Objection: this is just an instance of a well-known paradox, the 'paradox of the heap.' Parfit concedes that it is an argument of this sort, but says that in this case the most promising solution does not seem plausible. The most promising solution is to say that in some cases, it is just indeterminate whether some sand makes a heap. But it is hard to believe this about ourselves — it is hard to believe that it could be indeterminate whether some conscious being is me.

Here we face three options:

- (1) Accept the reductionist view, and say that in some of these cases it is indeterminate whether the person is me.
- (2) Accept the psychological criterion, and say that there is some sharp dividing line between these cases.

- (3) Say that all of these people would be me. This would seem to involve going for some version of the physical criterion.

Parfit suggests that most people will find (1) and (2) implausible, and opt for (3).

3.3 *The physical spectrum (§85)*

This kind of argument suggests to Williams that the physical criterion is a more promising approach to our identity over time.

Parfit replies that the very form of argument which can be used against the psychological criterion can be used against the physical criterion. He shows this using the example of the physical spectrum:

‘In a case close to the near end, scientists would replace 1% of the cells in my brain and body with exact duplicates. In the case in the middle of the spectrum, they would replace 50%. In a case near the far end, they would replace 99%, leaving only 1% of my original brain and body. At the far end, the ‘replacement’ would involve the complete destruction of my brain and body, and the creation out of new organic matter of a Replica of me. . . . There is here no physical continuity, since my brain and body are completely destroyed, and it is only later that the scientists create, out of new matter, my Replica.’

As with the psychological spectrum, we face three options:

- (1) Accept the reductionist view, and say that in some of these cases it is indeterminate whether the person is me.
- (2) Accept the physical criterion, and say that there is some sharp dividing line between these cases.
- (3) Say that all of these people would be me. This would seem to involve going for some version of the psychological criterion.

As above, Parfit suggests that most people will find (1) and (2) implausible, and opt for (3).

3.4 *The combined spectrum (§86)*

The problem is that it may seem difficult to square giving response (3) to both the psychological and the physical spectrum. Parfit suggests the possibility of a hybrid view, according to which either physical or psychological continuity is sufficient for identity.

But this hybrid view faces the combined spectrum:

‘At the near end of this spectrum is the normal case in which a future person would be fully continuous with me as I am now, both physically and psychologically. This person would be me in just the way that, in my actual life, it will be me who wakes up tomorrow. At the far end of this spectrum the resulting person would have no continuity with me as I am now, either physically or psychologically. In this case the scientists would destroy my brain and body, and then create, out of new organic matter, a perfect Replica of someone else. Let us suppose this person to be, not Napoleon, but Greta Garbo. We can suppose that, when Garbo was 30, a group of scientists recorded the states of all the cells in her brain and body.’

Again we face three options:

- (1) Accept the reductionist view, and say that in some of these cases it is indeterminate whether the person is me.
- (2) Say that there is some sharp dividing line between two of these cases.
- (3) Say that all of these people would be me.

The problem, as Parfit points out, is that option (3) in this case is absurd:

‘In considering the Combined Spectrum, we cannot accept this conclusion. In the case at the far end, the scientists destroy my brain and body, and then make, out of new matter, a Replica of Greta Garbo. There would be no connection, of any kind, between me and this resulting person. It could not be clearer that, in this case, the resulting person would not be me. We are forced to choose between the other two alternatives.’

This forces us to go for (1) or (2). Parfit concedes that option (2) would make sense if we believed ourselves to be immaterial souls, or other ‘separately existing entities.’ But he has already argued against this view.

Could we accept (2) without believing ourselves to be immaterial souls, and just say that there is some point in the combined spectrum at which the resulting person ceases to be me? About this kind of view Parfit says,

‘This view hardly differs from the Reductionist View. If we do draw such a line, we cannot believe that it has, intrinsically, either rational or moral significance. We must pick some point on this Spectrum, up to which we will call the resulting person me, and beyond which we will call him someone else. Our choice of this point will have to be arbitrary. We must draw this line between two neighbouring cases, though the difference between them is, in itself, trivial. If this is what we do, this should not affect our attitude towards these two cases. It would be clearly irrational for me to

regard the first case as being as good as ordinary survival, while regarding the second case as being as bad as ordinary death. When I consider this range of cases, I naturally ask, 'Will the resulting person be me?' By drawing our line, we have chosen to give an answer to this question. But, since our choice was arbitrary, it cannot justify any claim about what matters. And this is the most important claim in the Reductionist View. Our view differs only trivially from this view. Reductionists claim that, in some cases, questions about personal identity are indeterminate. We add the claim that, in such cases, we ought to give these questions answers, even if we have to do so in a way that is arbitrary, and that deprives our answers of any significance. I regard this view as one version of Reductionism, the tidy-minded version that abolishes indeterminacy with uninteresting stipulative definitions. Since the difference is so slight, I shall ignore this version of this view.'

The best response to these cases, Parfit thinks, is to go for option (1). In some cases, the question about whether some future conscious person will be you is an empty question, in the same way that questions about whether some future sports team will be the same team as some actually existing team is empty.