Paradoxes of personal identity: teletransportation, split brains, and immaterial souls

What is a theory of personal identity?

Suppose that you have a person, A, who exists at some time, and a person, B, who exists at some later time. A theory of personal identity is a theory which tries to answer the question: what does it take for A and B to be the **same person**?

Note that this is not a theory which described how we usually recognize and identify people. For example, suppose that I usually recognize you by some combination of your appearance and where in the room you sit. But of course two of you could, in an elaborate prank, change seats and have extensive reconstructive surgery which made each of you look much like the other. But this would not mean that either of you **became** the other person. You would not, for example, now be morally responsible for the actions your co-prankster performed yesterday.

One natural view of personal identity begins with the following view about the nature of persons:

Materialism about human beings

We are material (physical) objects.

On this view, we are certain material objects - namely, our bodies. This view is natural, because it fits with many things that we are inclined to say about ourselves. For example, we say that we have a certain weight and height, and are in a specific place; but what could occupy a place, and have a weight and height, other than a physical thing?

On this view, we are certain material objects - namely, our bodies. This view is natural, because it fits with many things that we are inclined to say about ourselves. For example, we say that we have a certain weight and height, and are in a specific place; but what could occupy a place, and have a weight and height, other than a physical thing?

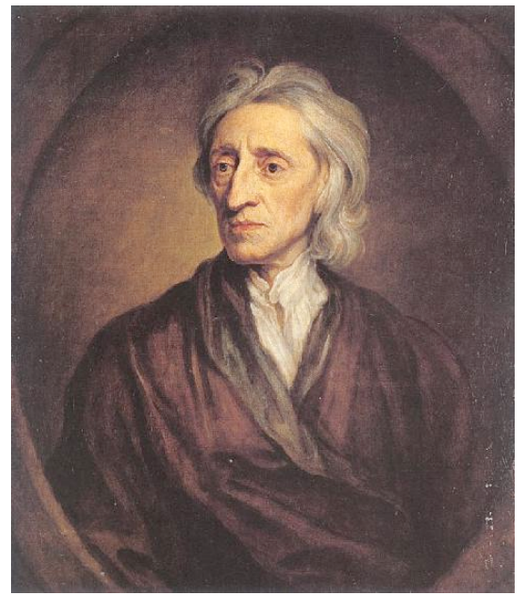
Materialism about human beings

We are material (physical) objects.

As we saw last time, the notion of a material object is not unproblematic; the example of the Ship of Theseus, for instance, raises problems for the idea that material objects can continue to exist over time. But there are other problems with the idea that persons, in particular, are material objects.

One important example is brought out by John Locke's example of the prince and the cobbler:

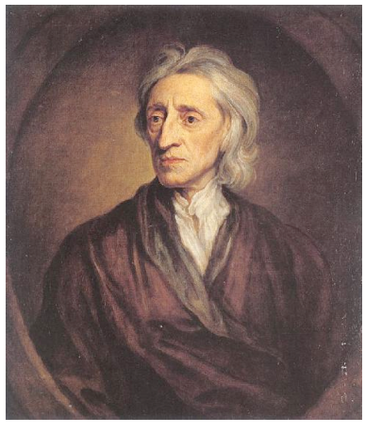
> For should the Soul of a Prince, carrying with it the consciousness of the Prince's past Life, enter and inform the Body of a Cobler as soon as deserted by his own Soul, every one sees, he would be the same Person with the Prince, accountable only for the Prince's Actions



What sort of example is Locke imagining here?

This seems to be a problem for materialist views of human persons. If Locke is right, and we can coherently imagine cases in which two persons "swap bodies", then it seems that we cannot be identical to our bodies. The case Locke imagines seems to be one in which a single organism is first one person, and then later becomes another person. But if this really is possible, it seems, materialism must be false.

> For should the Soul of a Prince, carrying with it the consciousness of the Prince's past Life, enter and inform the Body of a Cobler as soon as deserted by his own Soul, every one sees, he would be the same Person with the Prince, accountable only for the Prince's Actions



The sort of case Locke imagines is not just a problem for materialist theories of persons; it also suggests another theory of the nature of persons. Why, in this sort of case, do we all think that the person corresponding to the cobbler-body would be the prince? The key seems to be the fact that this person would have the "consciousness of the Prince's past life."

This suggests that what is essential to personal identity is not material continuity, **but rather some sort of continuity of consciousness**. This is the central idea of a competing theory of personal identity, which is sometimes called the **psychological theory** or the **memory theory** of personal identity. Locke is usually regarded as the first to defend this sort of theory.

This theory might be expressed as follows:

The memory theory of persons

If x and y are persons, then x=y if and only if x has memories of y (or vice versa).

The memory theory of persons

If x and y are persons, then x=y if and only if x has memories of y (or vice versa).

As Locke was aware, this theory has some surprising consequences. Here is one sort of problem that Locke raised for his own theory:

> § 22. But is not a Man Drunk and Sober the same Person, why else is he punish'd for the Fact he commits when Drunk, though he be never afterwards conscious of it? Just as much the same Person, as a Man that walks, and does other things in his sleep, is the same Person, and is answerable for any mischief he shall do in it.

What is the problem here? How should Locke respond?

The memory theory of persons

If x and y are persons, then x=y if and only if x has memories of y (or vice versa).

But the problems faced by the memory theory go well beyond these sorts of surprising consequences. As Thomas Reid, a Scottish contemporary of Locke, argued, certain sorts of examples seem to show that the theory leads to paradox.

> Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life: Suppose also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.
>
> These things being supposed, it follows, from Mr Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging—therefore, according to Mr Locke's doctrine, he is not the person who was flogged. Therefore, the general is, and at the same time is not the same person with him who was flogged at school.

The memory theory of persons

If x and y are persons, then x=y if and only if x has memories of y (or vice versa).

> Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life: Suppose also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.
>
> These things being supposed, it follows, from Mr Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging—therefore, according to Mr Locke's doctrine, he is not the person who was flogged. Therefore, the general is, and at the same time is not the same person with him who was flogged at school.

As with the example of the Ship of Theseus, it will be useful to introduce some names to bring out the sort of example Reid has in mind.

A = the boy at the time of the flogging
B = the officer at the time of the standard-taking
C = the general in "advanced life"

Then what Reid seems to be saying is that the following sort of scenario is possible:

> C has memories of the experiences of B, and B has memories of the experiences of A, but C does not have memories of the experiences of A.

The memory theory of persons

If x and y are persons, then x=y if and only if x has memories of y (or vice versa).

As with the example of the Ship of Theseus, it will be useful to introduce some names to bring out the sort of example Reid has in mind.

A = the boy at the time of the flogging
B = the officer at the time of the standard-taking
C = the general in “advanced life”

Then what Reid seems to be saying is that the following sort of scenario is possible:

> C has memories of the experiences of B, and B has memories of the experiences of A, but C does not have memories of the experiences of A.

We can see why this sort of scenario is problematic for the memory theory by laying out the following argument against that theory:

Reid’s paradox

1. x and y are the same person if and only if if the later has memories of the earlier. (The Memory Theory)
2. C has memories of the experiences of B.
3. C=B (1,2)
4. B has memories of the experiences of A.
5. B=A (1,4)
6. C does not have memories of the experiences of A.
7. C≠A (1,6)
8. C=A (3,5)

C. C=A & C≠A (7,8)

The memory theory of persons

If x and y are persons, then x=y if and only if x has memories of y (or vice versa).

Reid's paradox

1. x and y are the same person if and only if if the later has memories of the earlier. (The Memory Theory)
2. C has memories of the experiences of B.
3. C=B (1,2)
4. B has memories of the experiences of A.
5. B=A (1,4)
6. C does not have memories of the experiences of A.
7. C≠A (1,6)
8. C=A (3,5)

C. C=A & C≠A (7,8)

Reid's argument is a powerful one. It assumes only the transitivity of identity and the possibility of the sort of scenario described above. It is extremely difficult to deny that such scenarios are, in fact, possible.

So let's suppose that Reid's argument shows that the memory theory of persons, as stated above, is false. To respond to this argument, then, it seems that a proponent of that theory should try to find a way to reformulate her theory in such a way that it avoids Reid's objection.

One way to do this is to grant Reid that sometimes x can have no memories of y, and yet it still be the case that x=y. One might **still** hold that whenever x has memories of y, x=y.

The memory theory of persons

If x and y are persons, then x=y if and only if x has memories of y (or vice versa).

One way to do this is to grant Reid that sometimes x can have no memories of y, and yet it still be the case that x=y. One might **still** hold that whenever x has memories of y, x=y.

The memory theory, as formulated above, says:

x=y **if and only if** x has memories of y (or vice versa).

This can be thought of as the conjunction of the following two claims:

If x=y, then x has memories of y (or vice versa).

If x has memories of y (or vice versa), then x=y.

One idea is for the memory theorist to abandon the first of these claims, since Reid's example shows that the general might be the same person as the boy being flogged, despite the fact that neither has memories of the other. But the memory theory might still endorse the second of these claims.

The memory theory of persons

If x and y are persons, then x=y if and only if x has memories of y (or vice versa).

One way to do this is to grant Reid that sometimes x can have no memories of y, and yet it still be the case that x=y. One might **still** hold that whenever x has memories of y, x=y.

The memory theory, as formulated above, says:

x=y **if and only if** x has memories of y (or vice versa).

This can be thought of as the conjunction of the following two claims:

If x=y, then x has memories of y (or vice versa).

If x has memories of y (or vice versa), then x=y.

One idea is for the memory theorist to abandon the first of these claims, since Reid's example shows that the general might be the same person as the boy being flogged, despite the fact that neither has memories of the other. But the memory theory might still endorse the second of these claims.

However, this claim does not tell us everything we might want a theory of personal identity to tell us. It does not, in particular, seem to tell us exactly when x and y are the same person; if x does not have memories of y and y does not have memories of x, then this claim is simply silent on the question of whether x is y. But this might not seem very satisfactory; shouldn't a theory of persons explain what it takes, in any case, for x to be the same person as y?

The memory theory of persons

If x and y are persons, then x=y if and only if x has memories of y (or vice versa).

One way to do this is to grant Reid that sometimes x can have no memories of y, and yet it still be the case that x=y. One might **still** hold that whenever x has memories of y, x=y.

If x has memories of y (or vice versa), then x=y.

However, this claim does not tell us everything we might want a theory of personal identity to tell us. It does not, in particular, seem to tell us exactly when x and y are the same person; if x does not have memories of y and y does not have memories of x, then this claim is simply silent on the question of whether x is y. But this might not seem very satisfactory; shouldn't a theory of persons explain what it takes, in any case, for x to be the same person as y?

We can resolve this problem by reminding ourselves that identity is transitive: If A and B are the same person, and B and C are the same person, then A and C are the same person.

But, given the transitivity of identity, we know from the above claim that even if x does not have memories of y, x *must* be the same person as y if there is someone of whom x has memories and that person also shares memories with y.

This suggests the following version of the memory theory:

The modified memory theory of persons

x and y are the same person if and only if **either** (1) x has memories of y (or vice versa), **or** (2) there is some series of persons connecting x and y which is such that each person in the series has memories of the immediately preceding person in the series.

The modified memory theory of persons

x and y are the same person if and only if **either** (1) x has memories of y (or vice versa), **or** (2) there is some series of persons connecting x and y which is such that each person in the series has memories of the immediately preceding person in the series.

Reid's paradox

1. x and y are the same person if and only if if the later has memories of the earlier. (The Memory Theory)
2. C has memories of the experiences of B.
3. C=B (1,2)
4. B has memories of the experiences of A.
5. B=A (1,4)
6. C does not have memories of the experiences of A.
7. C≠A (1,6)
8. C=A (3,5)

C. C=A & C≠A (7,8)

It may be clear on an intuitive level how the modified memory theory avoids Reid's objection. But how, exactly, does the modified memory theory escape Reid's paradox?

The modified memory theory of persons

x and y are the same person if and only if **either** (1) x has memories of y (or vice versa), **or** (2) there is some series of persons connecting x and y which is such that each person in the series has memories of the immediately preceding person in the series.

This view is a bit complicated to keep in mind. It is perhaps easiest to think of the theory as breaking into two parts.

The memory requirement

If there are **no** memory relations -- whether direct or indirect -- between A and B, then A and B are not the same person.

The memory guarantee

If there **are** memory relations -- whether direct or indirect -- between A and B, then A and B are the same person.

It is a bit difficult to explain precisely what a “memory relation” is. But in the cases we will be interested in, a memory relation is a case of a person A having a conscious experience, and a later person B’s having a memory of having that conscious experience; in all the cases we are interested in, the memory is caused by the conscious experience it is a memory of. (The talk of “indirect” memory relations is introduced to handle cases like the one discussed by Reid.)

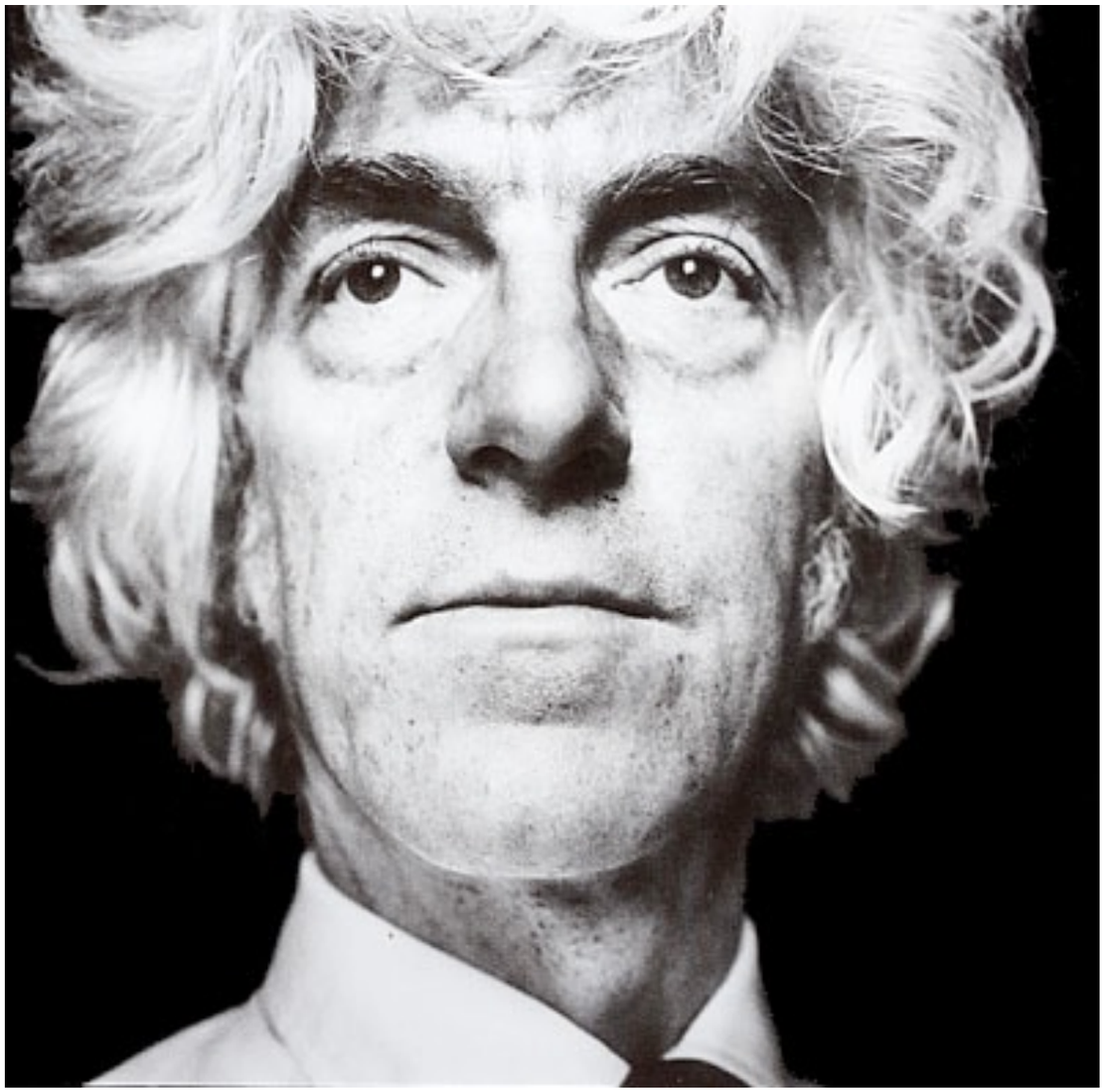
The memory requirement

If there are **no** memory relations -- whether direct or indirect -- between A and B, then A and B are not the same person.

The memory guarantee

If there **are** memory relations -- whether direct or indirect -- between A and B, then A and B are the same person.

A second sort of problematic case for the memory theory focuses not on the memory requirement, but the memory guarantee. This is Parfit's example of the teletransporter.



The memory requirement

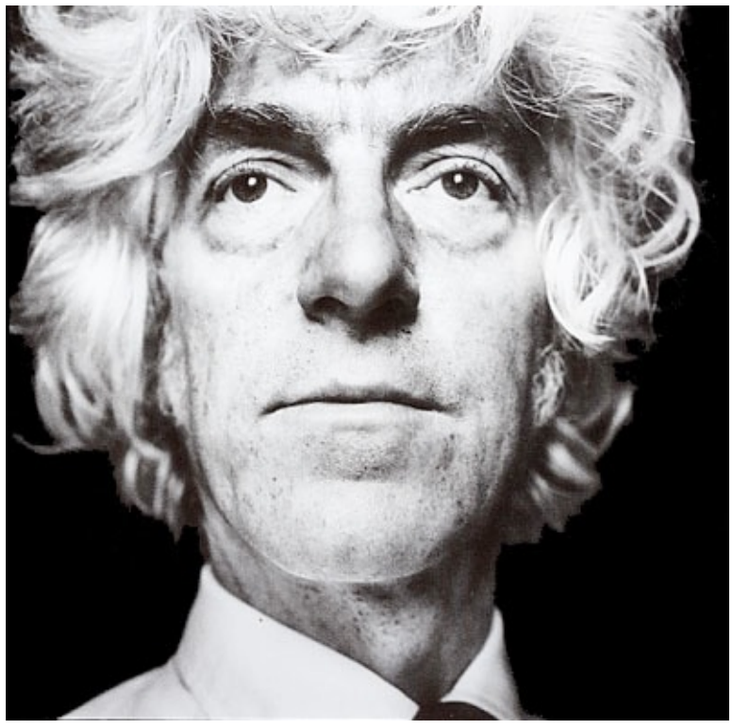
If there are **no** memory relations -- whether direct or indirect -- between A and B, then A and B are not the same person.

The memory guarantee

If there **are** memory relations -- whether direct or indirect -- between A and B, then A and B are the same person.

A second sort of problematic case for the memory theory focuses not on the memory requirement, but the memory guarantee. This is Parfit's example of the teletransporter.

The initial version of the journey by teletransportation to Mars seems relatively unproblematic, even if currently technologically impossible.



I enter the Teletransporter. I have been to Mars before, but only by the old method, a space-ship journey taking several weeks. This machine will send me at the speed of light. I merely have to press the green button. Like others, I am nervous. Will it work? I remind myself what I have been told to expect. When I press the button, I shall lose consciousness, and then wake up at what seems a moment later. In fact I shall have been unconscious for about an hour. The Scanner here on Earth will destroy my brain and body, while recording the exact states of all of my cells. It will then transmit this information by radio. Travelling at the speed of light, the message will take three minutes to reach the Replicator on Mars. This will then create, out of new matter, a brain and body exactly like mine. It will be in this body that I shall wake up.

Though I believe that this is what will happen, I still hesitate. But then I remember seeing my wife grin when, at breakfast today, I revealed my nervousness. As she reminded me, she has been often teletransported, and there is nothing wrong with *her*. I press the button. As predicted, I lose and seem at once to regain consciousness, but in a different cubicle. Examining my new body, I find no change at all. Even the cut on my upper lip, from this morning's shave, is still there.

The memory requirement

If there are **no** memory relations -- whether direct or indirect -- between A and B, then A and B are not the same person.

The memory guarantee

If there **are** memory relations -- whether direct or indirect -- between A and B, then A and B are the same person.

The problems begin with the arrival of the New Scanner.

The problems posed by this case are closely related to the problems posed by Reid's example. As in the case of Reid's argument, it will be useful to introduce some terms so that we can talk about this case clearly.

Original-Parfit = Parfit before he stepped into the teletransporter.

Earth-Parfit = the person who gets out of the teletransporter on earth.

Mars-Parfit = the person who gets out of the teletransporter on Mars.

Several years pass, during which I am often Teletransported. I am now back in the cubicle, ready for another trip to Mars. But this time, when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I leave the cubicle, and say to the attendant: 'It's not working. What did I do wrong?'

'It's working', he replies, handing me a printed card. This reads: 'The New Scanner records your blueprint without destroying your brain and body. We hope that you will welcome the opportunities which this technical advance offers.'

The attendant tells me that I am one of the first people to use the New Scanner. He adds that, if I stay for an hour, I can use the Intercom to see and talk to myself on Mars.

'Wait a minute', I reply, 'If I'm here I can't *also* be on Mars'.

Someone politely coughs, a white-coated man who asks to speak to me in private. We go to his office, where he tells me to sit down, and pauses. Then he says: 'I'm afraid that we're having problems with the New Scanner. It records your blueprint just as accurately, as you will see when you talk to yourself on Mars. But it seems to be damaging the cardiac systems which it scans. Judging from the results so far, though you will be quite healthy on Mars, here on Earth you must expect cardiac failure within the next few days.'

The attendant later calls me to the Intercom. On the screen I see myself just as I do in the mirror every morning. But there are two differences. On the screen I am not left-right reversed. And, while I stand here speechless, I can see and hear myself, in the studio on Mars, starting to speak.

The memory requirement

If there are **no** memory relations -- whether direct or indirect -- between A and B, then A and B are not the same person.

The memory guarantee

If there **are** memory relations -- whether direct or indirect -- between A and B, then A and B are the same person.

Original-Parfit = Parfit before he stepped into the teletransporter.

Earth-Parfit = the person who gets out of the teletransporter on earth.

Mars-Parfit = the person who gets out of the teletransporter on Mars.

The character in the story seems to be correct when he says “If I’m here I can’t also be on Mars.” But that is just another way of saying this:

Earth-Parfit ≠ Mars-Parfit

The problem is that both Earth-Parfit and Mars-Parfit stand in direct memory relations to Original-Parfit. Hence, if the memory guarantee is true, we know that each of the following must be true.

Several years pass, during which I am often Teletransported. I am now back in the cubicle, ready for another trip to Mars. But this time, when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I leave the cubicle, and say to the attendant: ‘It’s not working. What did I do wrong?’

‘It’s working’, he replies, handing me a printed card. This reads: ‘The New Scanner records your blueprint without destroying your brain and body. We hope that you will welcome the opportunities which this technical advance offers.’

The attendant tells me that I am one of the first people to use the New Scanner. He adds that, if I stay for an hour, I can use the Intercom to see and talk to myself on Mars.

‘Wait a minute’, I reply, ‘If I’m here I can’t *also* be on Mars’.

Someone politely coughs, a white-coated man who asks to speak to me in private. We go to his office, where he tells me to sit down, and pauses. Then he says: ‘I’m afraid that we’re having problems with the New Scanner. It records your blueprint just as accurately, as you will see when you talk to yourself on Mars. But it seems to be damaging the cardiac systems which it scans. Judging from the results so far, though you will be quite healthy on Mars, here on Earth you must expect cardiac failure within the next few days.’

The attendant later calls me to the Intercom. On the screen I see myself just as I do in the mirror every morning. But there are two differences. On the screen I am not left-right reversed. And, while I stand here speechless, I can see and hear myself, in the studio on Mars, starting to speak.

The memory requirement

If there are **no** memory relations -- whether direct or indirect -- between A and B, then A and B are not the same person.

The memory guarantee

If there **are** memory relations -- whether direct or indirect -- between A and B, then A and B are the same person.

The character in the story seems to be correct when he says “If I’m here I can’t also be on Mars.” But that is just another way of saying this:

Earth-Parfit ≠ Mars-Parfit

The problem is that both Earth-Parfit and Mars-Parfit stand in direct memory relations to Original-Parfit. Hence, if the memory guarantee is true, we know that each of the following must be true.

Earth-Parfit = Original-Parfit

Mars-Parfit = Original-Parfit

But, for reasons which are by now familiar, the three claims in blue cannot all be true: this trio of claims is inconsistent. So, if the scenario Parfit describes is really possible, it looks as though the memory guarantee implies a contradiction. But then the memory guarantee must be false.

Several years pass, during which I am often Teletransported. I am now back in the cubicle, ready for another trip to Mars. But this time, when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I leave the cubicle, and say to the attendant: ‘It’s not working. What did I do wrong?’

‘It’s working’, he replies, handing me a printed card. This reads: ‘The New Scanner records your blueprint without destroying your brain and body. We hope that you will welcome the opportunities which this technical advance offers.’

The attendant tells me that I am one of the first people to use the New Scanner. He adds that, if I stay for an hour, I can use the Intercom to see and talk to myself on Mars.

‘Wait a minute’, I reply, ‘If I’m here I can’t *also* be on Mars’.

Someone politely coughs, a white-coated man who asks to speak to me in private. We go to his office, where he tells me to sit down, and pauses. Then he says: ‘I’m afraid that we’re having problems with the New Scanner. It records your blueprint just as accurately, as you will see when you talk to yourself on Mars. But it seems to be damaging the cardiac systems which it scans. Judging from the results so far, though you will be quite healthy on Mars, here on Earth you must expect cardiac failure within the next few days.’

The attendant later calls me to the Intercom. On the screen I see myself just as I do in the mirror every morning. But there are two differences. On the screen I am not left-right reversed. And, while I stand here speechless, I can see and hear myself, in the studio on Mars, starting to speak.

The memory requirement

If there are **no** memory relations -- whether direct or indirect -- between A and B, then A and B are not the same person.

The memory guarantee

If there **are** memory relations -- whether direct or indirect -- between A and B, then A and B are the same person.

Earth-Parfit ≠ Mars-Parfit

Earth-Parfit = Original-Parfit

Mars-Parfit = Original-Parfit

But, for reasons which are by now familiar, the three claims in blue cannot all be true: this trio of claims is inconsistent. So, if the scenario Parfit describes is really possible, it looks as though the memory guarantee implies a contradiction. But then the memory guarantee must be false.

We can use Parfit's example of the teletransporter and the New Scanner to generate the following paradox:

1. Lockean body switching is possible.
2. If Lockean body switching is possible, then teletransportation is possible.
3. If teletransportation is possible, then teletransportation using the New Scanner is possible

---

C. Teletransportation using the New Scanner is possible. (1,2,3)

And we've already seen that we have a convincing argument that the conclusion of this argument is false.

This might suggest that we should re-think our view of Lockean body switching, and revisit the possibility that materialism about persons is true.

However, one can generate problems for materialism quite similar to the problems to which teletransportation gives rise for the memory theory.

However, one can generate problems for materialism quite similar to the problems to which teletransportation gives rise for the memory theory.

These are cases of **fission**. Suppose that instead of Parfit stepping into a teletransporter, he decided to undergo an ambitious new form of surgery.

In this surgery, one's body is sawn in half. The left half is then joined with a perfect replica of the right half, and the right half is then joined with a perfect replica of the left half.

Let's call the resultant persons Left-Parfit and Right-Parfit. It is obvious that Left-Parfit ≠ Right-Parfit. But it seems that if materialism is true, Left-Parfit = Original-Parfit and Right-Parfit=Original Parfit. After all, each of Left- and Right-Parfit are physically connected to Original-Parfit.

Might the materialist reply that neither of Left- and Right-Parfit have **enough** of a connection to Original-Parfit? Perhaps one must, from moment to moment, have **more than 50%** of the cells of someone in order to be identical to them.

But this sort of view is open to at least three objections.

Let’s call the resultant persons Left-Parfit and Right-Parfit. It is obvious that Left-Parfit ≠ Right-Parfit. But it seems that if materialism is true, Left-Parfit = Original-Parfit and Right-Parfit=Original Parfit. After all, each of Left- and Right-Parfit are physically connected to Original-Parfit.

Might the materialist reply that neither of Left- and Right-Parfit have **enough** of a connection to Original-Parfit? Perhaps one must, from moment to moment, have **more than 50%** of the cells of someone in order to be identical to them.

But this sort of view is open to at least three objections.

1. As Parfit says, it is hard to believe that there could be a single “cut off point.” Suppose that the surgeon accidentally includes a bit more of Original-Parfit in the left half. Could that really determine whether Original-Parfit survives the surgery?
2. Moreover, it seems a bit like cheating, since we would not find the “>50%” requirement plausible if the other half did not survive. Suppose that more than half of someone’s body was destroyed in a terrible accident. Wouldn’t we think that it was great if medical science were able to save the person’s life by replicating the destroyed portion of the body and re-joining it to the surviving portion?
3. One might reply to these worries by saying that it is not the whole body which determines personal identity, but rather just some part of the body - like the brain. But even here one might worry about the seeming possibility of partial brain transplants. Suppose that we acquired the ability to cure brain cancer by replicating the cancerous portion of the brain, removing the cancerous part, and replacing it with the replica. Would that really kill the patient? Would it matter exactly what % of the brain had to be removed? What would be the cut-off point?

At this stage, both materialism and the memory theory might seem to be in pretty bad shape.

Parfit suggests a radical response to these problems. According to Parfit, when we talk about “personal identity” or “being the same person”, we aren’t really talking about an all-or-nothing thing. Rather, we are just talking about degrees of psychological similarity. So when I say that A and B are the same person, what I really mean is just: A and B are psychologically connected in certain interesting ways.

One useful comparison (which Parfit suggests elsewhere) is a comparison of persons to clubs, or teams. Suppose that we begin a personal identity discussion club at Notre Dame. People gradually leave and join the club, and some of the rules change, and eventually people decide that at meetings things other than personal identity may occasionally be discussed. At one of the meetings (in 2048) someone says: “Is this really the **same club** as the one formed way back in 2014?”

Parfit suggests, and this seems right, that this is not a very deep question. The club in 2048 is similar in some ways to our club, and different in other ways; there is no **further fact** about whether the two clubs are **really the same.** We could decide to say that they are identical or distinct, but our choice seems somewhat arbitrary.

Parfit’s radical suggestion is that people are, in this way, like clubs. When we ask, “Is Original-Parfit really the same person as Mars-Parfit, or Earth-Parfit?” we are not asking a very deep question. Each is similar in certain important ways to Original-Parfit, and that is pretty much the end of the story. There is simply no further, fundamental fact about which one is identical to Original-Parfit.

This view has some surprising consequences. One is that questions about death and survival also do not have all-or-nothing answers. Think about Earth-Parfit after he comes out of the New Scanner. One naturally thinks that he should be very upset about the fact that he is going to die soon. But, if Parfit is right, he should be much consoled by the fact that Mars-Parfit, who is psychologically extremely similar to him, will continue to live -- after all, ordinary survival just is a matter of there being someone psychologically quite similar to me who continues to exist. (Compare the survival of a club.)

Parfit's radical suggestion is that people are, in this way, like clubs. When we ask, "Is Original-Parfit really the same person as Mars-Parfit, or Earth-Parfit?" we are not asking a very deep question. Each is similar in certain important ways to Original-Parfit, and that is pretty much the end of the story. There is simply no further, fundamental fact about which one is identical to Original-Parfit.

This view has some surprising consequences. One is that questions about death and survival also do not have all-or-nothing answers. Think about Earth-Parfit after he comes out of the New Scanner. One naturally thinks that he should be very upset about the fact that he is going to die soon. But, if Parfit is right, he should be much consoled by the fact that Mars-Parfit, who is psychologically extremely similar to him, will continue to live -- after all, ordinary survival just is a matter of there being someone psychologically quite similar to me who continues to exist. (Compare the survival of a club.)

One might think that Earth-Parfit could protest:

> "But Mars-Parfit isn't me! Why should I feel better about dying because this other guy will live!"

But if Parfit is right, this is just confused. Mars-Parfit sort of is Earth-Parfit — they are psychologically similar in important ways and, if Parfit is right, that is all there is to personal identity.

Some connections between this view and four-dimensionalist theories of change.

Is there any way to avoid the paradoxical consequences to which materialism and the memory theory lead, while **not** joining Parfit in abandoning the natural view that survival of persons is always an all-or-nothing matter?

There is. One could adopt a **dualist theory** of personal identity, according to which persons are immaterial souls. Then survival is always an all-or-nothing matter: it is just a matter of the continued existence of a soul. So, strictly speaking, people, not being material things, do not have weights and heights; but they are closely connected to bodies, which of course do have weights and heights.

This view has some advantages. Assuming that immaterial souls are indivisible, the problems of division illustrated by the examples of fission and teletransportation cannot be used against the dualist. (Of course, dualism doesn't say exactly what does happen in these cases - just that the original person survives if and only if one the post-surgery (or post-teletransportation) bodies is attached to his soul. But, souls being invisible, it might be quite hard to tell.)

But reflection on the nature of the relationship between soul and body can make this view seem difficult to accept.

It is very plausible that what happens to your body can affect your mental life, and that mental events also have physical effects. The dualist who agrees with these points is an **interactionist dualist**, since she believes in genuine causal interactions between souls and the material world.

But this can seem mysterious; how could an immaterial thing, which lacks physical attributes like mass and momentum, bring about effects in the physical world?

One worry about this is that it seems that the interactionist dualist has to think that certain conservation laws involving physical quantities have exceptions. Consider, for example, the conservation of energy and the conservation of momentum. It would seem that to bring about effects the in physical world, a soul would have to bring about a change in the energy or momentum of some physical system. But wouldn't such a change violate conservation laws, since the system in question would be, for our purposes, physically isolated?

The question of whether the dualist can make sense of mind-body interactions is a difficult and important one, which deserves more discussion than we can give it here.

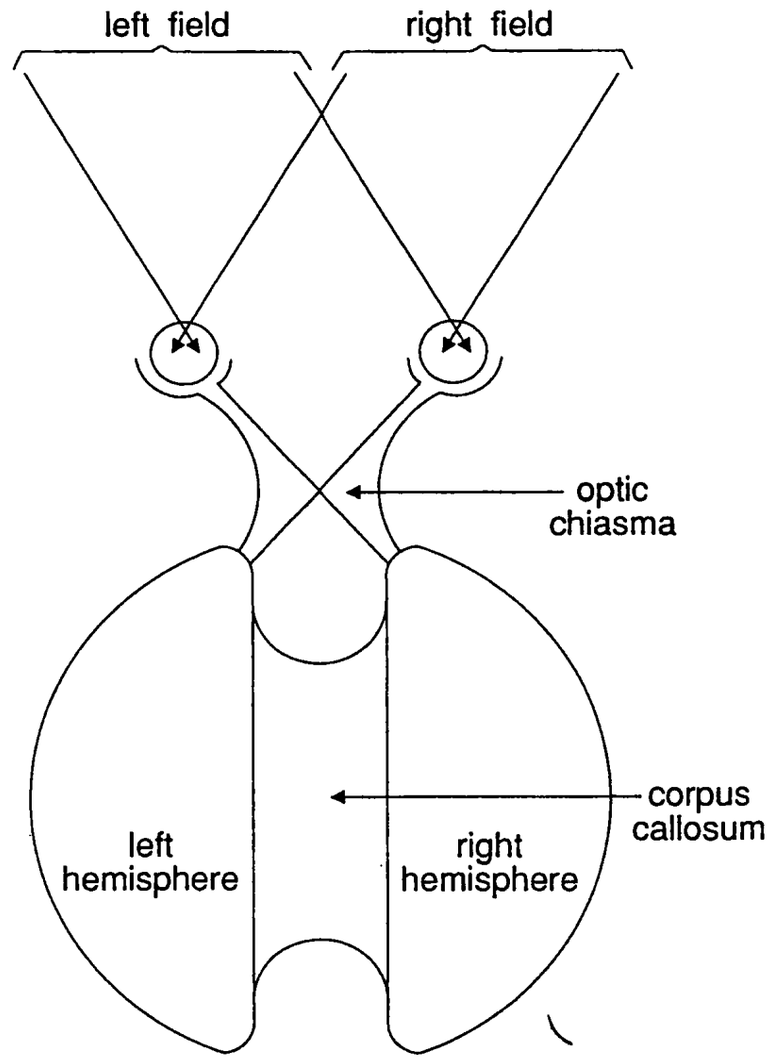
Fortunately (or unfortunately), a final paradox of personal identity seems to apply to the dualist just as well as to the materialist and psychological theorists. This arises from the cases of brain bisection discussed by Parfit (and by Nagel in the optional reading).

Fortunately (or unfortunately), a final paradox of personal identity seems to apply to the dualist just as well as to the materialist and psychological theorists. This arises from the cases of brain bisection discussed by Parfit (and by Nagel in the optional reading).

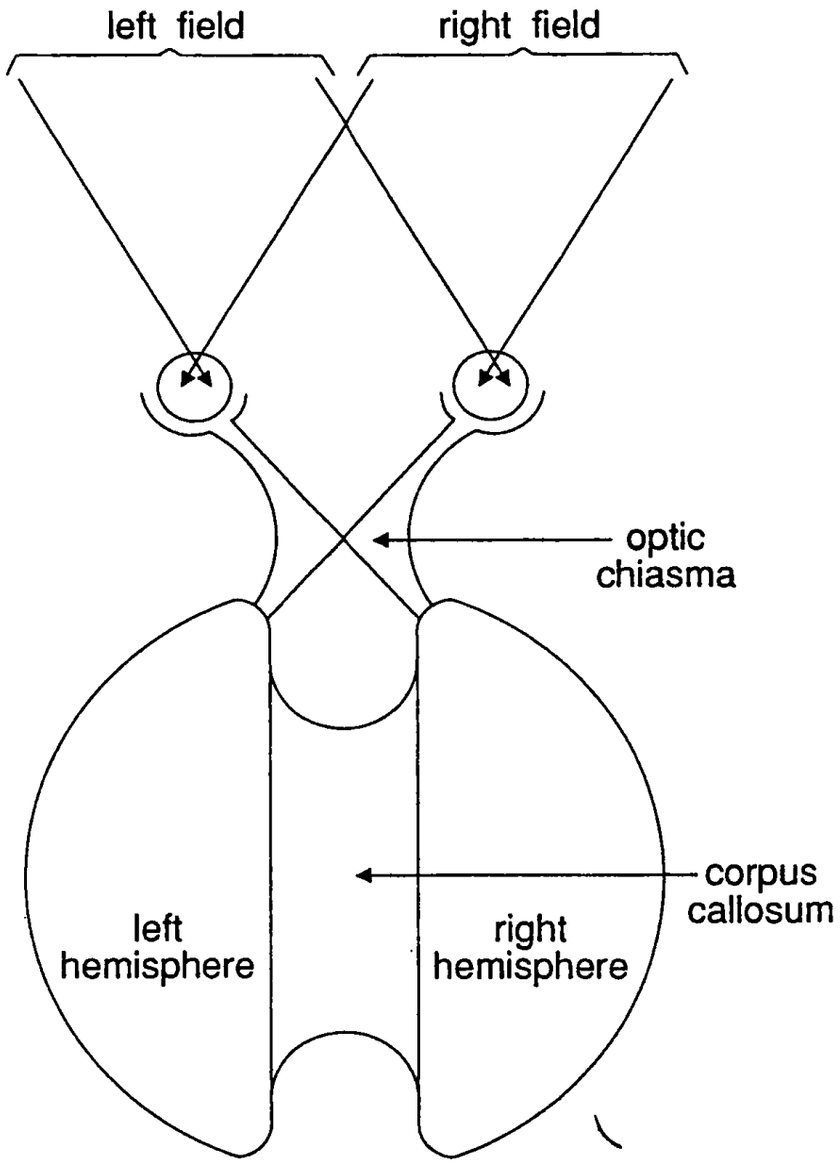
These are studies of patients whose corpus callosum has been severed. The corpus callosum is a pathway which connects the left and right hemispheres of the human brain and, in normal subjects, allows the two hemispheres of the brain to exchange information.

If the corpus callosum is severed, the two hemispheres of the brain cannot exchange information. So any sensory data about the environment available to, for example, the left hemisphere, will not be available to guide the movements of the left hand, which is controlled by the right hemisphere. Information available only to the right hemisphere will not be reportable in speech, since speech is controlled by the left hemisphere.

The results of giving sensory data to just one of the hemispheres of the brain of such a patient are striking.

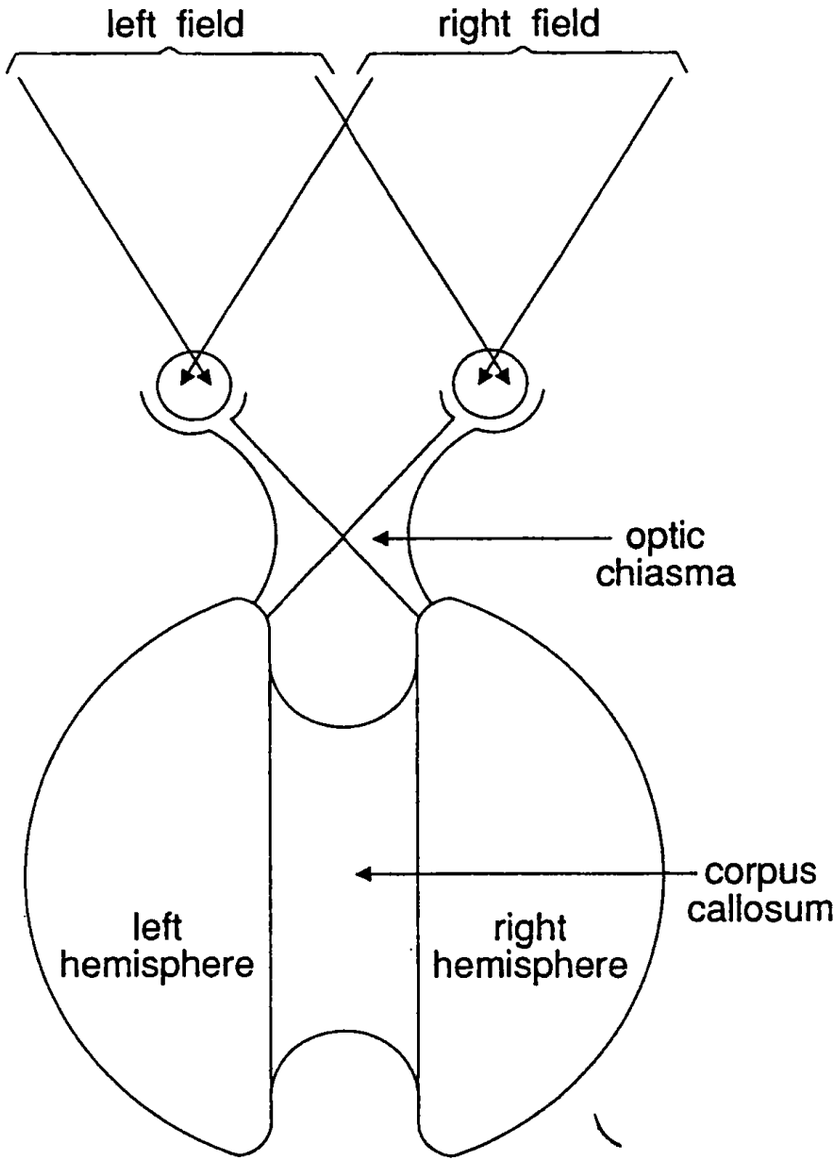


The results of giving sensory data to just one of the hemispheres of the brain of such a patient are striking. (The following quote is from Nagel's "Brain bisection and the unity of consciousness", which is linked from the course web site.)



The results are as follows. What is flashed to the right half of the visual field, or felt unseen by the right hand, can be reported verbally. What is flashed to the left half field or felt by the left hand cannot be reported, though if the word 'hat' is flashed on the left, the left hand will retrieve a hat from a group of concealed objects if the person is told to pick out what he has seen. At the same time he will insist verbally that he saw nothing. Or, if two different words are flashed to the two half fields (e.g. 'pencil' and 'toothbrush') and the individual is told to retrieve the corresponding object from beneath a screen, with both hands, then the hands will search the collection of objects independently, the right hand picking up the pencil and discarding it while the left hand searches for it, and the left hand similarly rejecting the toothbrush which the right had lights upon with satisfaction.

The results of giving sensory data to just one of the hemispheres of the brain of such a patient are striking. (The following quote is from Nagel's "Brain bisection and the unity of consciousness", which is linked from the course web site.)



> One particularly poignant example of conflict between the hemispheres is as follows. A pipe is placed out of sight in the patient's left hand, and he is then asked to write with his left hand what he was holding. Very laboriously and heavily, the left hand writes the letters P and I. Then suddenly the writing speeds up and becomes lighter, the I is converted to an E, and the word is completed as PENCIL. Evidently the left hemisphere has made a guess based on the appearance of the first two letters, and has interfered, with ipsilateral control. But then the right hemisphere takes over control of the hand again, heavily crosses out the letters ENCIL, and draws a crude picture of a pipe.[6]

Why do these split brain cases lead to paradox?

The following two principles seem quite plausible (especially if, like memory theorists, we think that the nature of persons is tied closely to consciousness):

Ownership

Every conscious experience must be an experience of someone.

Awareness

If someone has a conscious experience, it must be at least in principle possible for them to be aware of that experience.

Now think about a case in which a split-brain patient has a red stimulus presented to the right half of their visual field, and a blue stimulus presented to the left half of their visual field. If you ask the subject what color they see, they will say “Red”, since this was the color presented to the part of the eye which feeds input to the left hemisphere of the brain, which controls speech.

So it is clear that there is a conscious experience of red; so, by **Ownership**, there must be someone who is having this experience. Let’s call this person “Mr. Red.”

If you put a pen in the left hand of the left hand of the subject, and ask what color was just seen, that hand will write “Blue.” So it seems that there must have been a conscious experience of blue -- otherwise, how would the hand know what color to write?

But if there is a conscious experience of blue, by **Ownership** someone must have had this experience. Let us call the person who has this experience “Mr. Blue.”

Ownership

Every conscious experience must be an experience of someone.

Awareness

If someone has a conscious experience, it must be at least in principle possible for them to be aware of that experience.

So it is clear that there is a conscious experience of red; so, by **Ownership**, there must be someone who is having this experience. Let's call this person "Mr. Red."

But if there is a conscious experience of blue, by **Ownership** someone must have had this experience. Let us call the person who has this experience "Mr. Blue."

Now the crucial question is: Is Mr. Red the same person as Mr. Blue? It seems to follow from **Awareness** that they are not the same person. After all, if you ask Mr. Red whether he has had any experience of blue, he will say "No." And no amount of introspection on his part will allow him to remember having a conscious experience of this sort; and of course this is not because he forgot having the experience, but because he was never aware of having it. But then, by **Awareness**, he *didn't* have it.

Hence it seems that Mr. Red ≠ Mr. Blue. So there are two persons in the body of the split brain patient.

This is a bit weird on its own. But further oddities result from consideration of what this conclusion says about non-split-brain patients, like us.

Ownership

Every conscious experience must be an experience of someone.

Awareness

If someone has a conscious experience, it must be at least in principle possible for them to be aware of that experience.

So it is clear that there is a conscious experience of red; so, by **Ownership**, there must be someone who is having this experience. Let's call this person "Mr. Red."

But if there is a conscious experience of blue, by **Ownership** someone must have had this experience. Let us call the person who has this experience "Mr. Blue."

Hence it seems that Mr. Red ≠ Mr. Blue. So there are two persons in the body of the split brain patient.

This is a bit weird on its own. But further oddities result from consideration of what this conclusion says about non-split-brain patients, like us.

There seem to be three things we can say:

1. While the split brain patients are in experiments of this sort, there are two persons inhabiting their body; but, at other times, there is just one person inhabiting their body.

2. Split brain patients always have two persons inhabiting their body, but non-split brain subjects do not.

3. All of us, split-brain and non-split-brain subjects alike, have two (or more) persons inhabiting their body.

There seem to be three things we can say:

But each of these options seems, for various reasons, absurd.

1. While the split brain patients are in experiments of this sort, there are two persons inhabiting their body; but, at other times, there is just one person inhabiting their body.

If this were true, then simply flashing some red and blue lights at someone would bring a new person into existence; and turning off the lights would kill that person.

2. Split brain patients always have two persons inhabiting their body, but non-split brain subjects do not.

If this were true, then severing the corpus callosum of an epileptic patient would bring a new person into existence; and reversing the surgery would kill that person.

3. All of us, split-brain and non-split-brain subjects alike, have two (or more) persons inhabiting their body.

Non-split brain patients never have conscious experiences of which they are not aware; but then it would follow that there is a person inhabiting my body which never has any conscious experiences at all. But then in what sense does that person even exist?

One can, of course, follow Parfit and say that our talk about persons, or subjects of experience, is just a convenient fiction for talking about conscious experiences. The split-brain cases illustrate that there are cases in which this convenient fiction breaks down; in cases like the one described above, there is a red experience and a blue experience, and that is all that we can say; there is no further fact about whether these experiences are experiences of the same person, or not. This is just the surprising denial of the reality if persons which we were trying to avoid.