

Formulating  
the  
psychological  
theory

Spectrum  
arguments

Cases of fission

Does survival  
matter?

*Are you a psychological  
thing?*

Last time we were discussing 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 Cobbler 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

This kind of case poses a problem for the materialist's answer to the survival question.

**The survival question:** What does it take for for some person at some other time to be you?

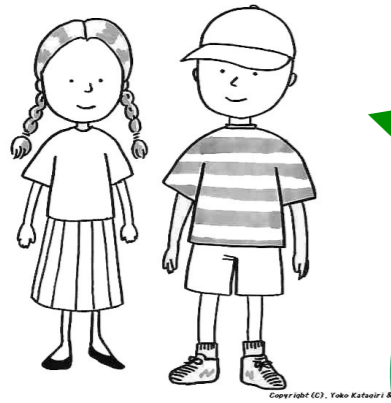
**The small changes theory of survival:** X is me just in case X is separated from me by a series of small physical changes over time.

**The survival question:** What does it take for for some person at some other time to be you?

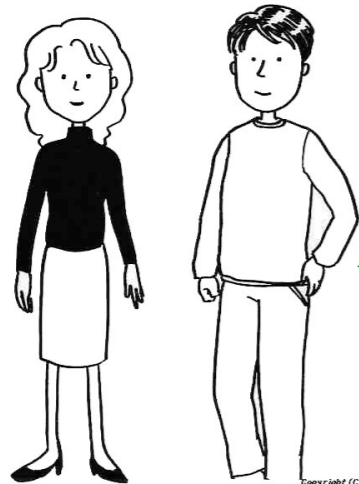
The kind of theory I want to talk about today gives an entirely different answer to this question. According to this theory, the key to survival is not **physical** connectedness, but **psychological** connectedness.

**The psychological theory of survival:** X is me just in case X has the right kinds of psychological connections to me

This was John Locke's theory. His view of personhood can be illustrated by considering a few different stages in the lives of some people.



psycho-  
logical  
connection



psycho-  
logical  
connection



This was John Locke's theory. His view of personhood can be illustrated by considering a few different stages in the lives of some people.

What makes the child, the adult, and the elderly person stages of the same person? The materialist says: because they are the same material thing. Locke thought: it is because of **psychological connections** between the individuals.

This gives us the wanted result that the prince before the switch, and the cobbler after the switch, are the same person.

This view has some notable advantages over the views we have discussed so far.

It captures the intuition that you could wake up in a different body than the one you now have.

It avoids the materialist's problem with explaining how you could be the same material thing despite changes in your parts over time. No more fears that haircuts might be the end of you!

It seems to avoid the materialist's problems with making sense of the possibility of life after death. For surely God could create a being which stood in all of the right psychological connections to you, even if your body has decayed.

If one is attracted to the psychological theory of survival, how should you answer the identity question?

If one is attracted to the psychological theory of survival, how should you answer the identity question?

**The identity question:** What are you? Are you an organism, an immaterial soul, or something else?

This question can seem tough for the psychological theorist to answer. For consider the prince and the cobbler case. It looks tempting to say that one person — call him Mr. X — was at one time identical to the prince's body, and at another time identical to the cobbler's body. But of course the prince's body is not identical to the cobbler's body. But doesn't that violate the transitivity of identity?

**The identity question:** What are you? Are you an organism, an immaterial soul, or something else?

The best view for the psychological theorist is to adopt a view of identity over time known as **four-dimensionalism**. According to this view, much as physical things are made up out of distinct spatial parts — like my left and right hand — objects that exist over a period of time are made out of distinct **temporal parts**.

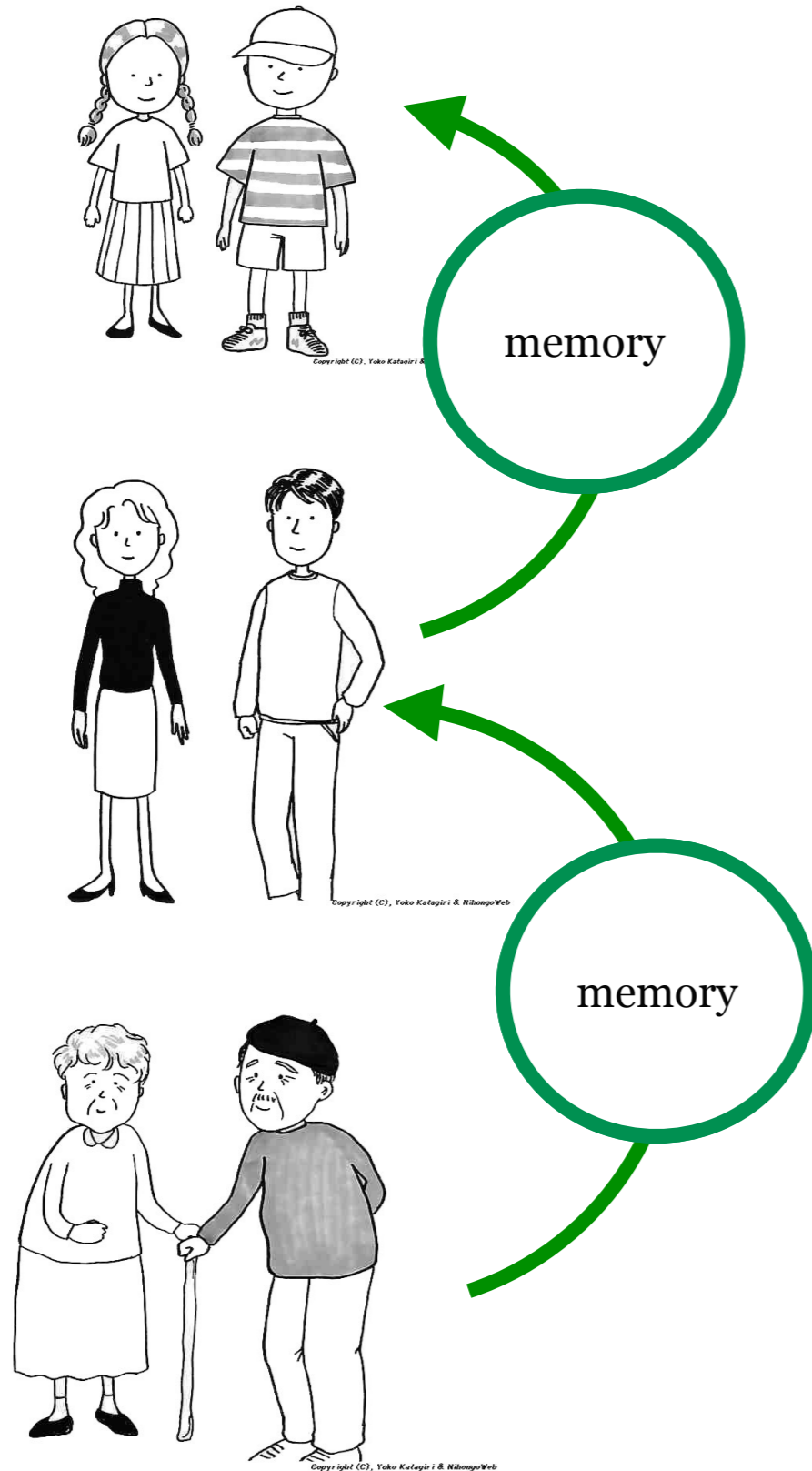
Just as objects extend through space by having parts in different spatial locations, so objects extend through time by having different temporal parts at different times. And what makes these these distinct temporal parts all parts of the same person is their standing in certain psychological relations to each other.

**The identity question:** What are you? Are you an organism, an immaterial soul, or something else?

So the psychological theorist's answer to the identity question can be put like this:

**Psychological identity:** I am a collection of temporal parts that stand in the right psychological relations





**Psychological identity:** I am a collection of temporal parts that stand in the right psychological relations

But what are the relevant psychological relations?

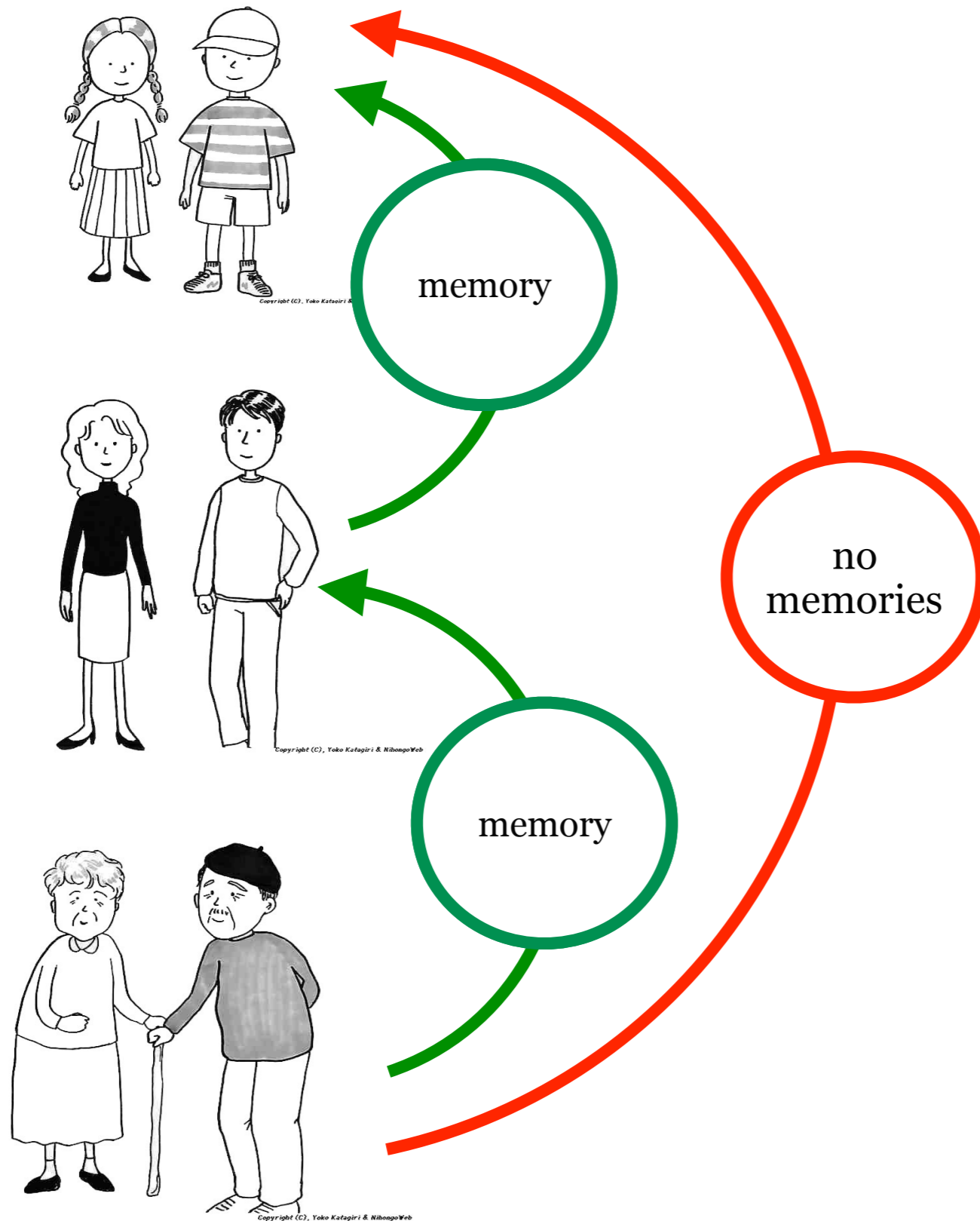
Locke's answer was: relations of **memory**.

But, as Locke's contemporary Thomas Reid noted, this answer leads to an immediate problem.

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.

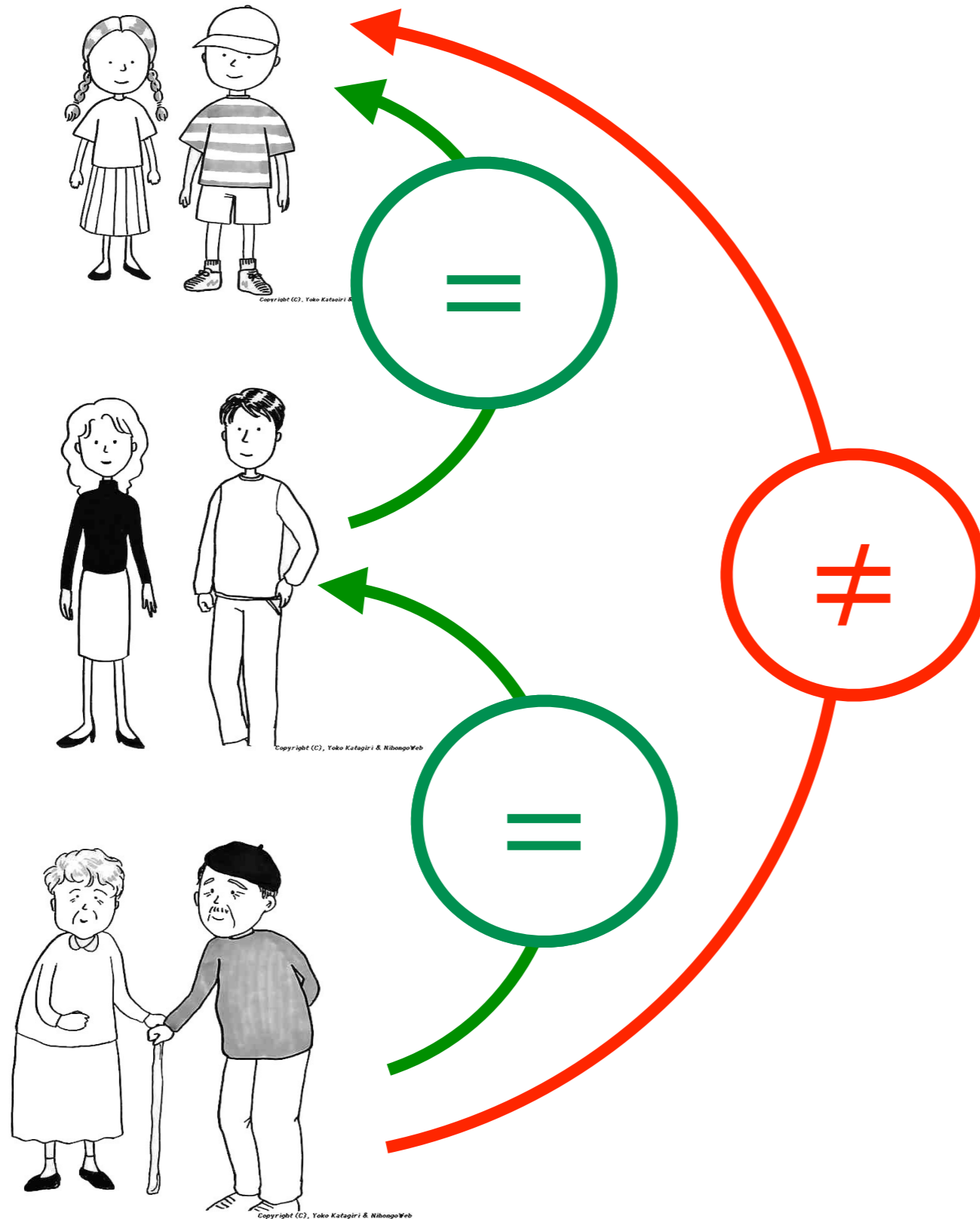
These things being supposed, it follows, from Mr Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows, if there be any truth in logic, that the general is the same person 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.





We can illustrate the kind of scenario that Reid had in mind.

This is problematic because, if identity of persons is determined by memory or its absence, Reid's example leads to an impossible constellation of identity facts.

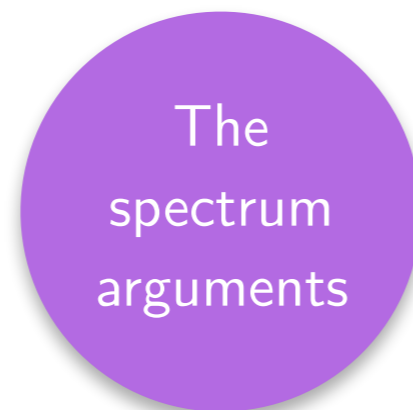


How should the psychological theorist reply?

One promising reply: introduce the notion of an **indirect memory relation**, which is related to memory relations in the way that **ancestor** is related to parent.

So far, the psychological theory looks to be in good shape.

Let's now look at two fundamental challenges to the view. As we'll see, these can either be regarded as objections to the psychological theory, or as challenges to some of our most deeply held preconceptions about what we are and what our survival means.



The  
spectrum  
arguments

We can introduce the spectrum arguments via a thought experiment Parfit discusses in the reading.

I am the prisoner of some callous neuro-surgeon, 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 ... 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? ...

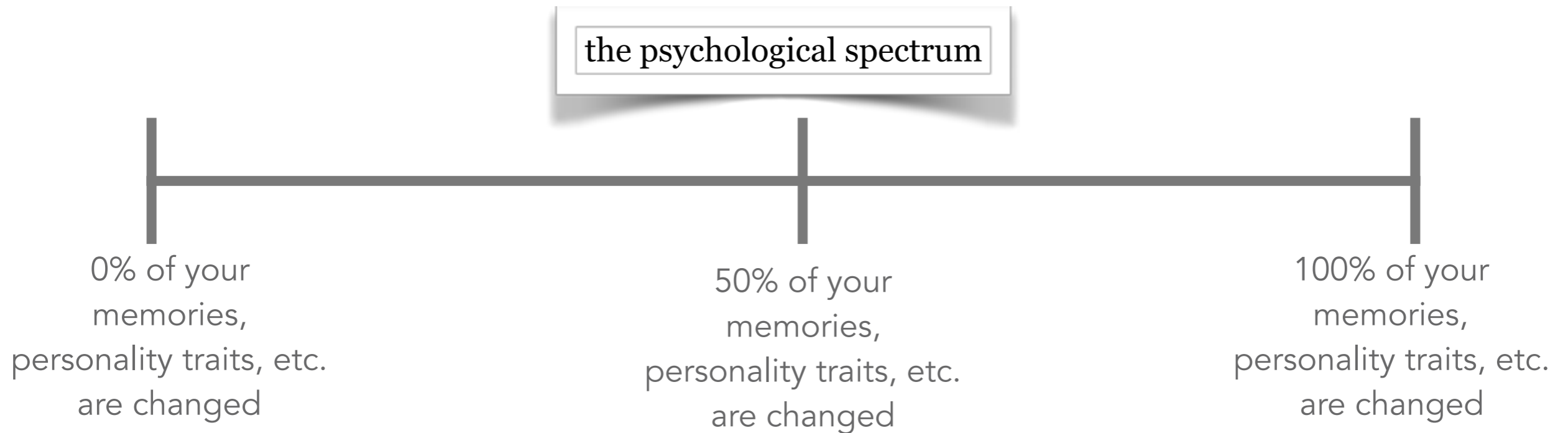
Surely not. ...

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. ... 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.

Intuitively, at the end of this series of unfortunate events, you would still be in pain. But what must the psychological theorist say about this case?

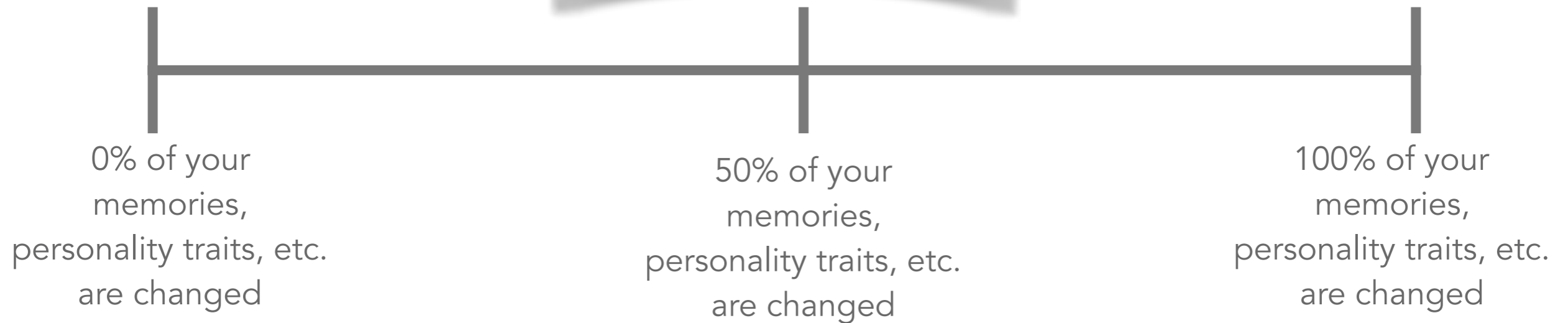
As Parfit says, we can think of these kinds of psychological changes as falling on a spectrum.



The intuitive response to the torture example implies that you survive even on the far right edge of the spectrum.

This seems to show that the psychological theory is false.

## the psychological spectrum

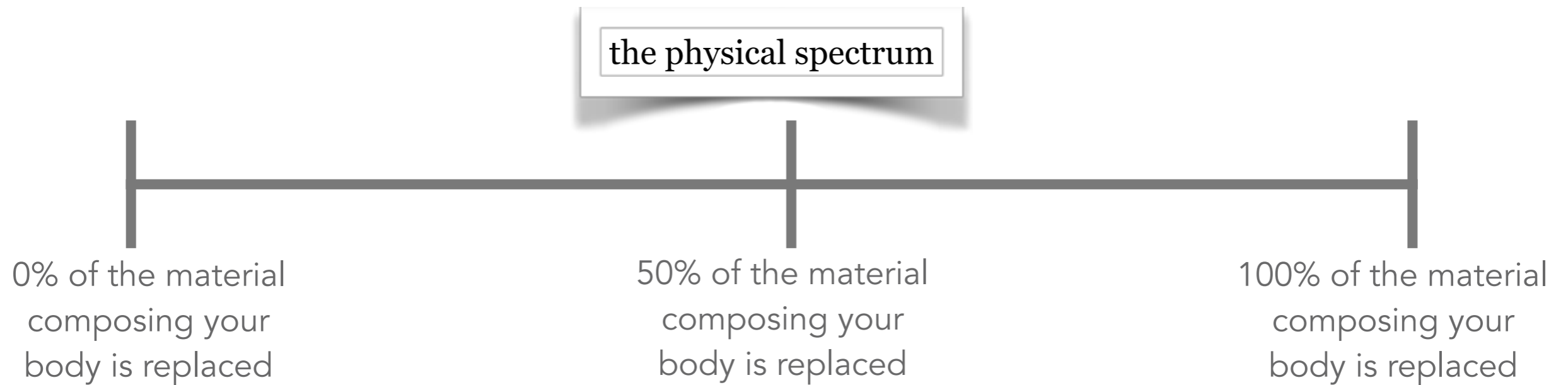


But even if you don't share this intuition, the psychological spectrum raises some tough questions for the psychological theory.

The psychological theorist seems forced to say that there is some particular line on the spectrum where the organism in question would cease to be you. But can that be right? Why should, say, losing 43.13% of one's personality be consistent with survival, but losing 43.14% enough for you to cease to exist?



So far this looks like good news for materialism. But, Parfit points out, we can consider a similar argument against the materialist.



In the middle of this spectrum, 50% of your body is replaced. At the far end, your body is destroyed and a duplicate put in its place.

## the physical spectrum

0% of the material  
composing your  
body is replaced

50% of the material  
composing your  
body is replaced

100% of the material  
composing your  
body is replaced

In the middle of this spectrum, 50% of your body is replaced. At the far end, your body is destroyed and a duplicate put in its place.

Here, the materialist seems forced to say that there is some special percentage of the matter in your body such that if that percentage is replaced, you would cease to exist. But again that seems quite difficult to believe; how could one cell make all of the difference?

Here's an analogy. Suppose that I am an impoverished philosophy professor, and definitely not rich.

Now suppose that a wealthy benefactor who loves philosophy decides to give me some money. But he does this in an eccentric way: by adding 1 cent to my bank account every second.

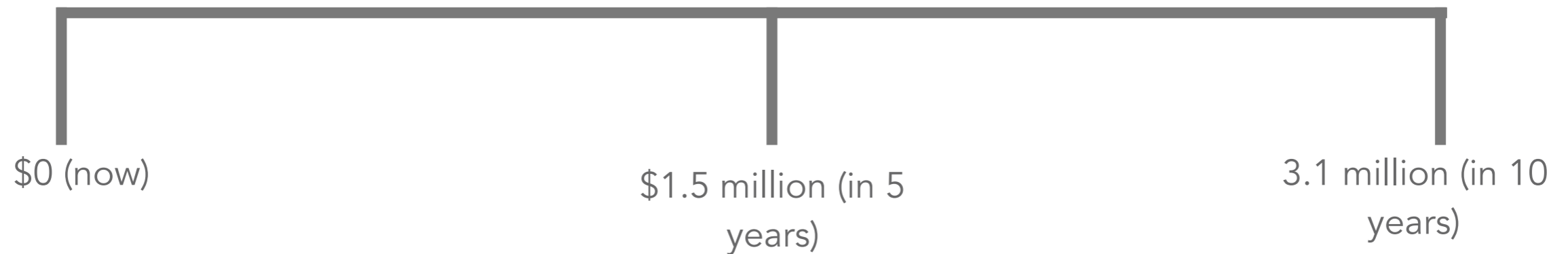
At the end of 10 years, I will have \$3.1 million in my bank account, and will be rich.

We can chart my progress using the wealth spectrum.

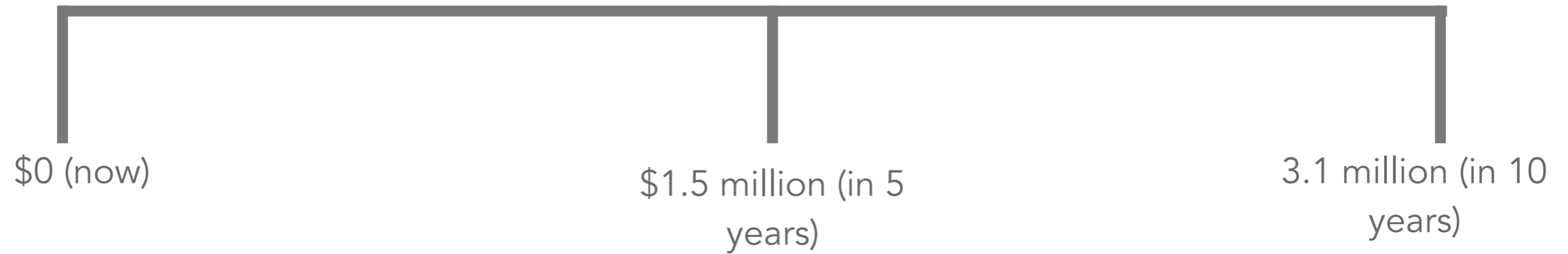
Now suppose that a wealthy benefactor who loves philosophy decides to give me some money. But he does this in an eccentric way: by adding 1 cent to my bank account every second.

At the end of 10 years, I will have \$3.1 million in my bank account, and will be rich.

We can chart my progress using the wealth spectrum.



When, precisely, did I become rich?



When, precisely, did I become rich?

It is overwhelmingly plausible that there is no precise answer to this question. I was not rich at the beginning, and I was rich at the end. But for a bunch of seconds in the middle, there is just no fact of the matter about whether I was rich. Being rich is, in this sense, not an “all or nothing” thing.

Parfit thinks that the psychological theorist should say that being you is like being rich. The psychological theorist should say that for some middle stages in the psychological spectrum, there is just no fact of the matter about whether that person is you.

On this kind of view, your existence is not an all or nothing matter.

So Parfit thinks that the moral of the spectrum arguments is not that the psychological theory is false, but that we should change a fundamental part of our view about what our own continued existence amounts to.

Writing about the psychological spectrum, Parfit says

“[One] assumes that, in each of these cases, the resulting person either would or would not be me. This is not so. The resulting person would be me in the first few cases. In the last case he would not be me. In many of the intervening cases, neither answer would be true. I can always ask, ‘Am I about to die? Will there be some person living who will be me?’ But, in the cases in the middle of this Spectrum, there is no answer to this question.”

Parfit argues that it is a consequence of both the psychological and the materialist answers to the survival question that sometimes the question of whether some future person will be me has no answer.

Parfit argues that it is a consequence of both the psychological and the materialist answers to the survival question then sometimes the question of whether some future person will be me has no answer.

Can you think of any way that the psychological theorist (or the materialist) could avoid this result? Is it a result that they should want to avoid?

The  
spectrum  
arguments

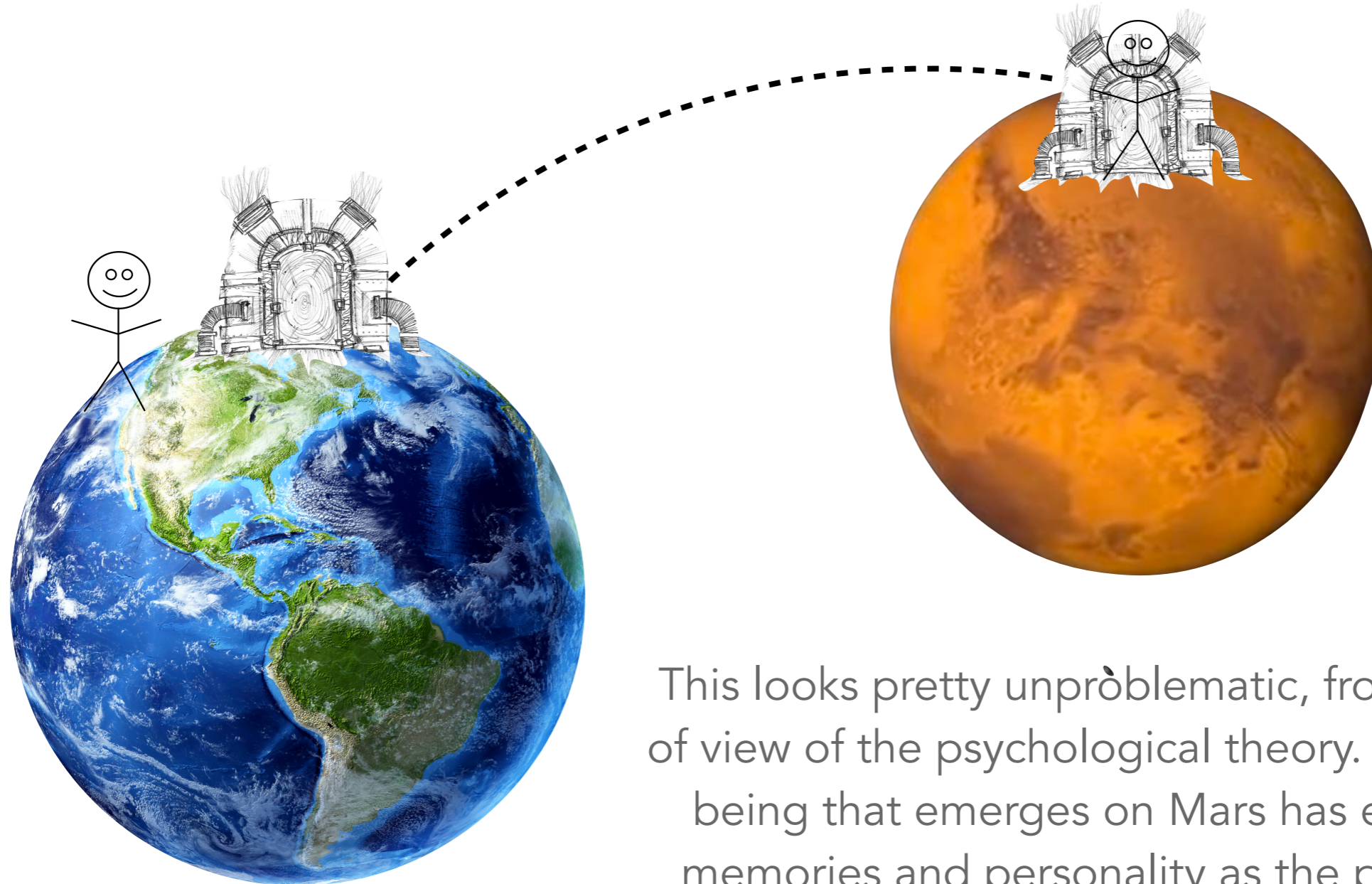
Cases of  
fission

Let's turn next to our second problem for the psychological theory: the problem raises by cases of **fission**. One good way of presenting these cases uses the possibility of teletransportation.



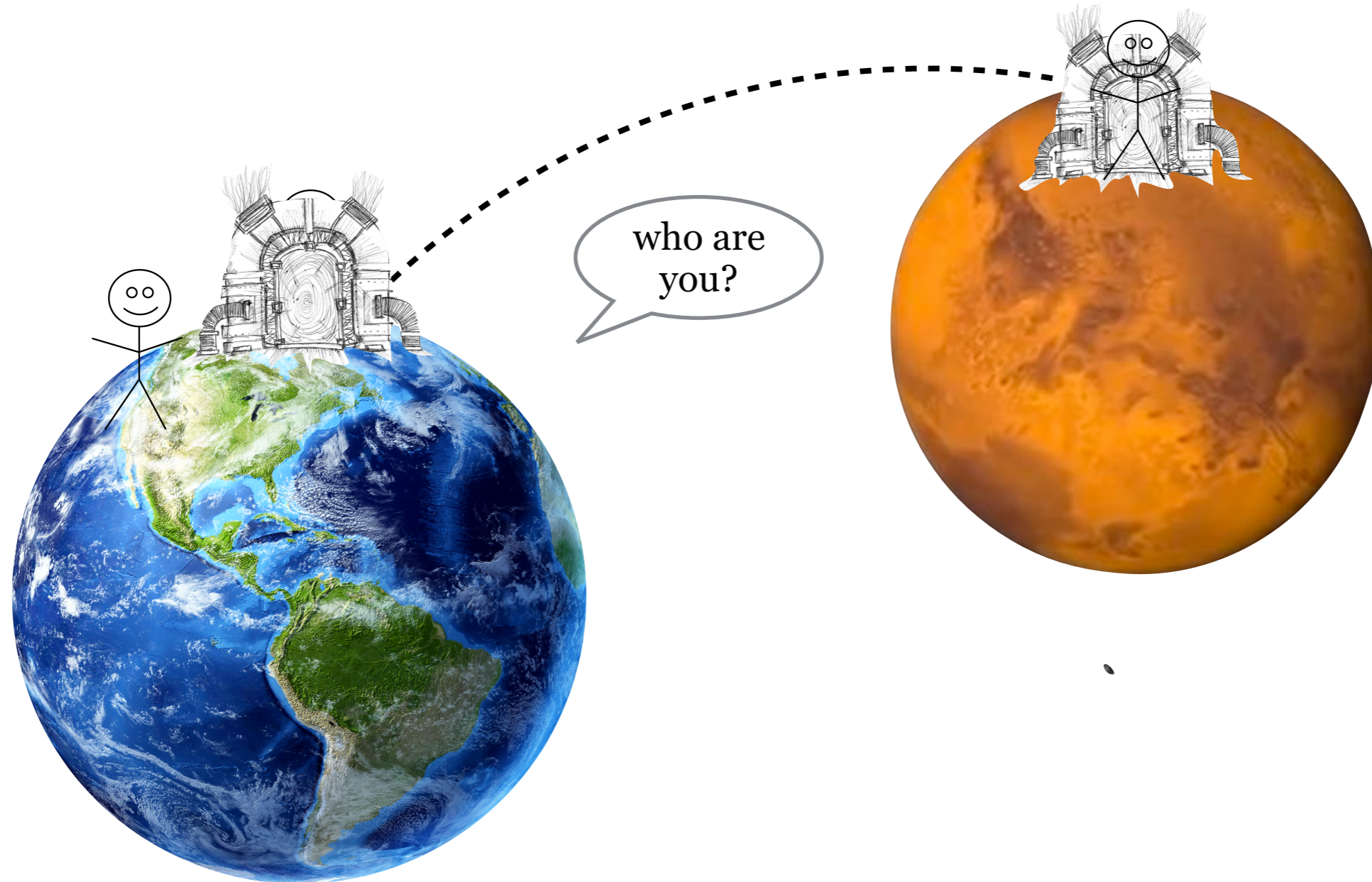


The teletransporter was invented as a way of traveling quickly from Earth to the now-colonized planet of Mars. One simply steps into the teletransporter on Earth, at which time all of the data about my cells is recorded and transmitted near the speed of light to Mars, at which time that data is used by the teletransporter there to reconstitute me.

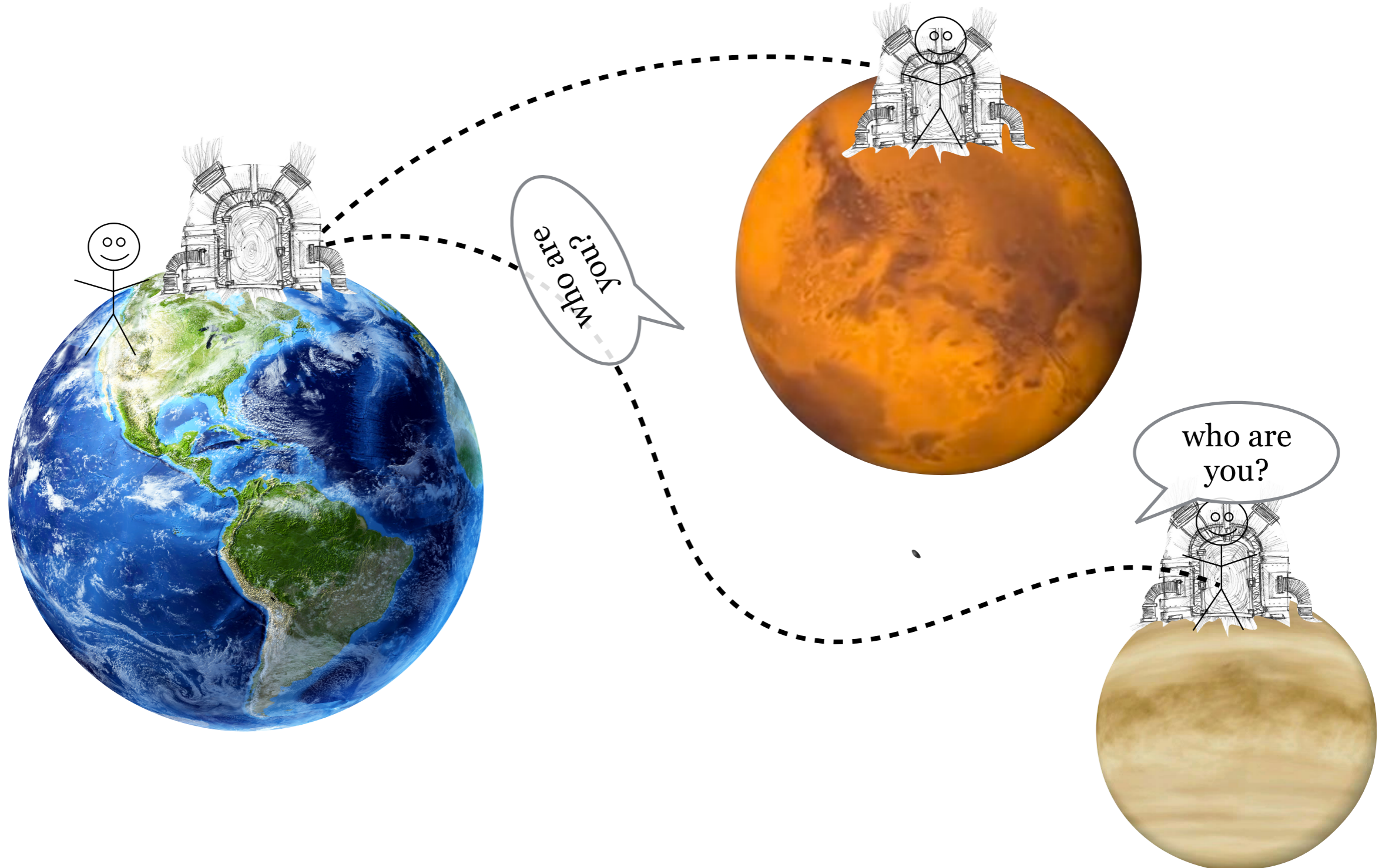


This looks pretty unproblematic, from the point of view of the psychological theory. After all, the being that emerges on Mars has exactly the memories and personality as the person who stepped into the machine on Earth. So it is the same person.

But problems are not far away. What happens if the teletransportation machine on earth, after copying all of the information about the cells of the person who steps into the teletransporter, simply leaves the body in the teletransportation machine untouched? This is what Parfit calls the 'New Scanner.'



Or we can imagine that there is another teletransportation machine located on the surface of Venus, to which the machine on earth simultaneously transmits the relevant cellular information.



We appear to face a problem which is in some ways similar to the problem posed by the Ship of Theseus. Let's focus on the version of the story in which there are teletransportation machines on Mars and Venus.

Let's call the person who steps into the teletransporter on Earth 'Earthy,' the one who steps out on Mars 'Marsy,' and the one who steps out on Venus 'Venusy.'

We have already seen that, if the psychological theory is true, then the idea that a single person can travel (and continue to exist!) via teletransportation is unproblematic. So we know that, if the psychological theory is true, then:

Earthy = Marsy

Earthy = Venusy

But the following seems clearly true:

Marsy ≠ Venusy

But, for reasons we have already discussed — namely, the fact that identity is transitive — these three claims do not sit well together. So it appears that the psychological theory implies a contradiction.

Earthy = Marsy  
Earthy = Venusy  
Marsy ≠ Venusy

Basically the same point could be made about the version of the story on which, after the transmission to Mars, the individual who steps into the teletransporter on Earth steps back out. To tell that version of the story, we'd just need to introduce two names — Earthy-1 and Earthy-2 — for the individual on earth pre-teletransportation, and the individual who exists after the teletransportation.

These cases look bad for the psychological theory of survival. One might think that they provide a reason to favor a materialist view. What would a materialist say about teletransportation?

These cases look bad for the psychological theory of survival. One might think that they provide a reason to favor a materialist view. What would a materialist say about teletransportation?

As it turns out, though, the materialist faces a quite similar problem. This problem can be introduced by describing an ambitious new form of surgery.

We are all familiar with surgeries in which parts of one's body are removed. One might remove a tumor, or an organ, or a limb.

Let us suppose that in the future medical technology continues to improve. It is now possible to amputate half of a person's body.

Fortunately, prosthetics have also improved, so that it is now possible to make an exact duplicate of the half that has been removed, and attach that to the original half.



It seems as though it should be, in principle, possible to survive this surgery. So the materialist, who thinks that you are identical to an organism, must also say that it is possible for an organism to survive this surgery.

So, it seems, they must endorse the following claim:

An organism can survive even if 50% of that organism's matter is replaced.

But this assumption leads to trouble.

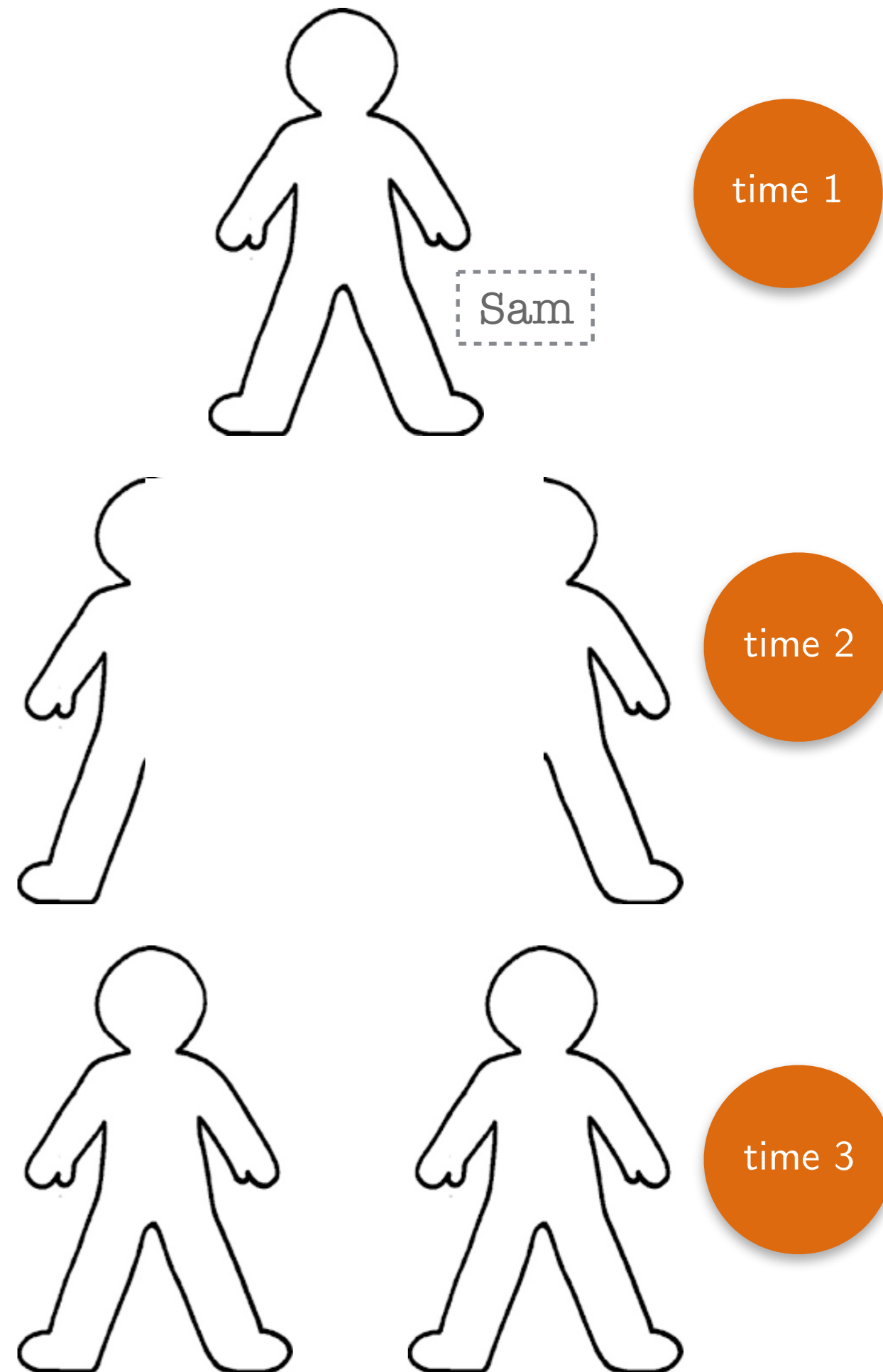


An organism can survive even if 50% of that organism's matter is replaced.

Suppose that we take a healthy patient, Sam, and cut him in half.

We then, as in the previous surgery, make duplicates of the two halves, and join them to the two severed halves.

Call the two resulting individuals "Lefty" and "Righty."





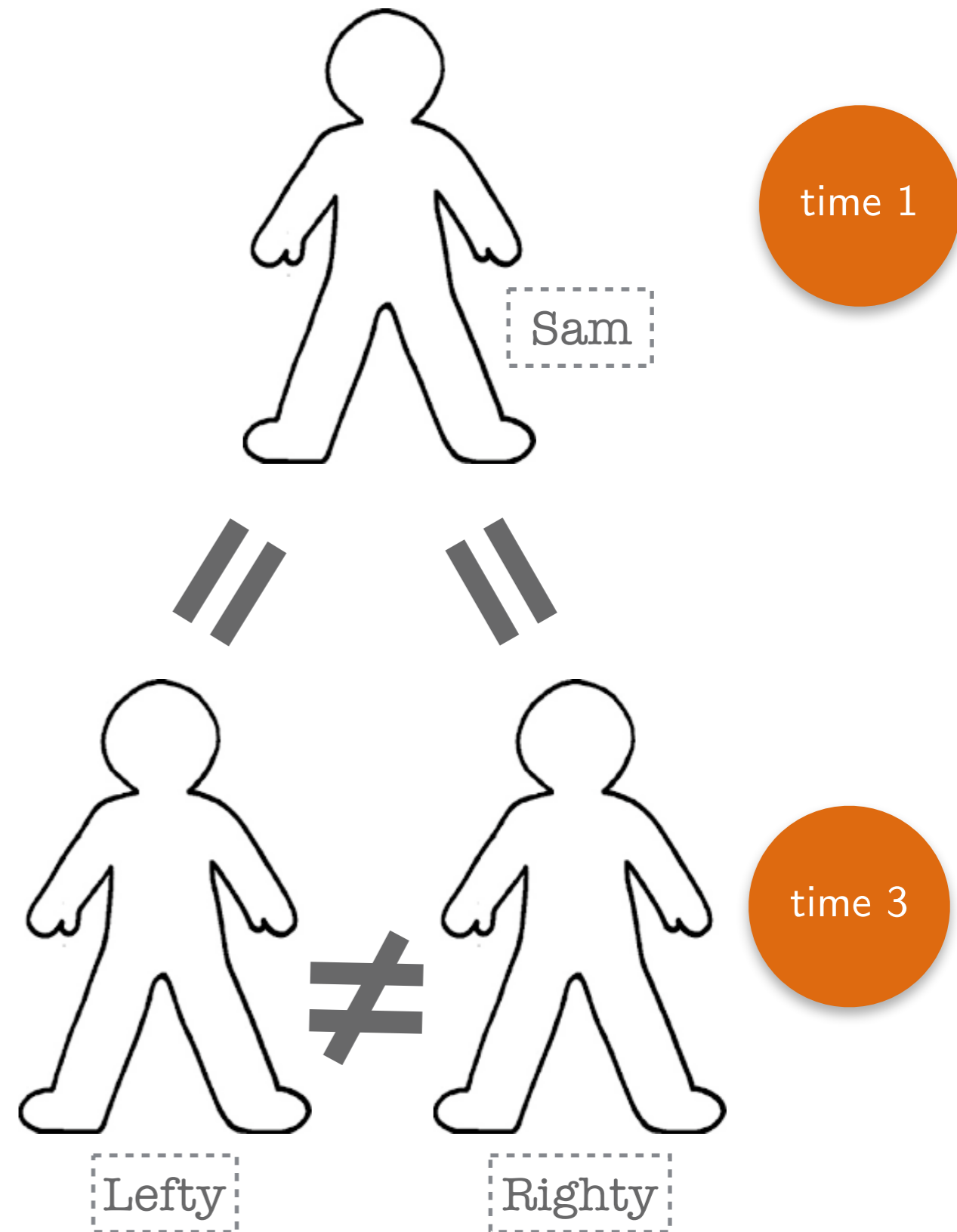
An organism can survive even if 50% of that organism's matter is replaced.

The problem is that the materialist who endorses the above principle seems forced to say that Sam=Lefty and Sam=Righty.

But, obviously, Lefty  $\neq$  Righty.

And this is a contradiction.

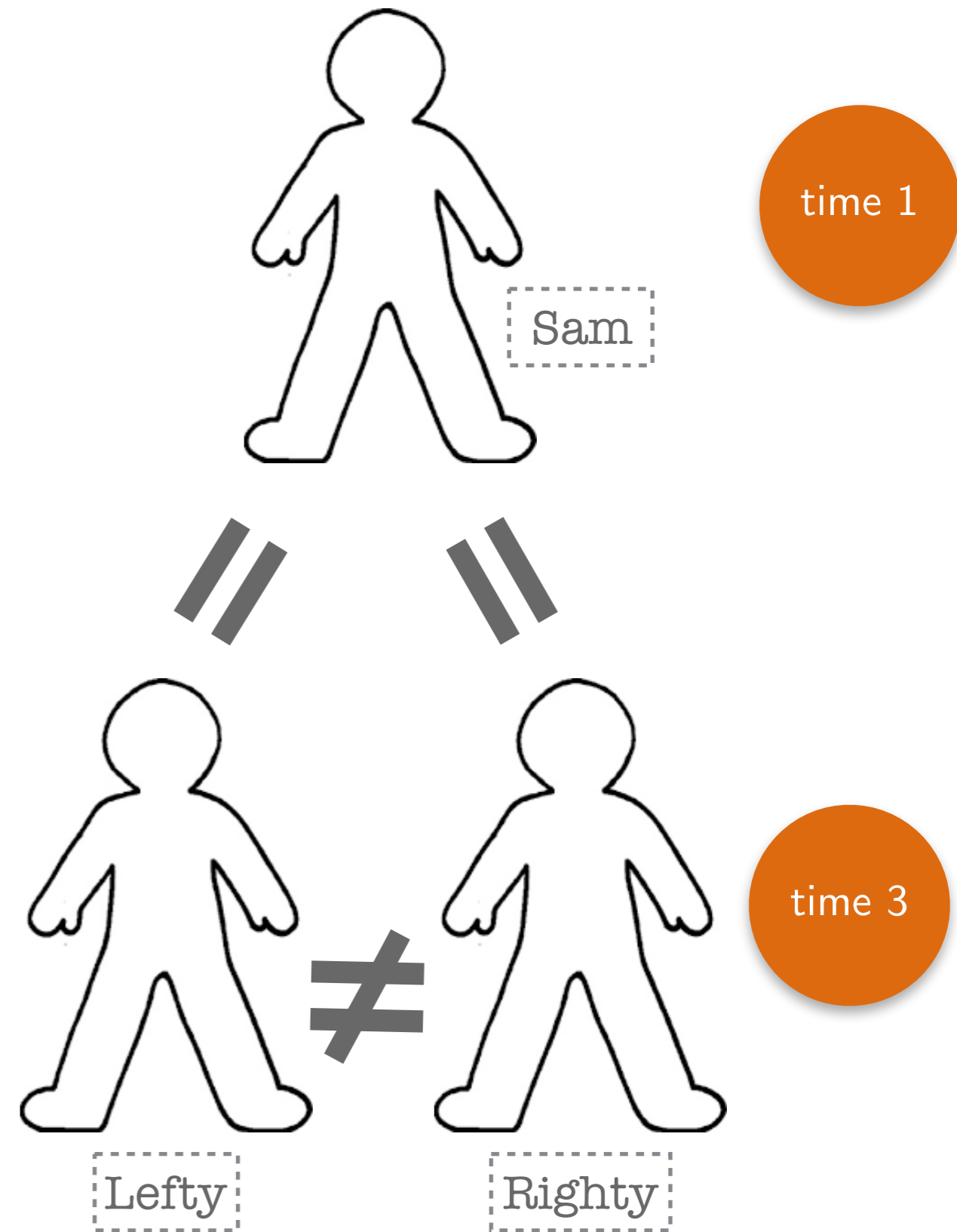
How should the materialist respond?



How should the materialist respond?

Is it plausible to say that 50.1% of the material of the body must be preserved?

Imagine that you had to undergo an extensive surgery of this kind, and that it was impossible to tell whether you had lost just over, or just less than 50% of your body. On this kind of proposal, you would then be unsure whether you were a new person with memories of someone else's actions, or the same person who went into the surgery. Would this be reasonable?



Cases of fission are among the most challenging problems for materialist and psychological theories of survival.

Parfit offers a response to these cases. But it has a startling consequence. On Parfit's view, there are some cases in which it makes no sense to care about whether one survives. As he puts it, **survival is not what matters.**

To illustrate what that means, let's look at one of the cases he discusses.

To illustrate what that means, let's look at one of the cases he discusses.

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.'

On Parfit's view, this is just about as good as ordinary survival. What do you think?

On Parfit's view, this is just about as good as ordinary survival. What do you think?

Most people, on first hearing, think that Parfit's views are crazy. And Parfit knows this. Let's look at why Parfit has this view, starting with a return to a case we've already discussed in connection with materialism.

### **The brain transplant**

While driving, you have gotten into a car accident, and your body has been irreparably damaged, but your brain has not. Fortunately (for you), your passenger's body was unharmed, but she has sustained irreparable damage to her brain. Surgeons quickly transplant your brain into your passenger's body.

It is natural to think that you could survive this surgery.

Now let's look at a tougher case which Parfit discusses.

*My Division.* My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine.

Let's call the resulting people Lefty and Righty. Then we seem to have four options:

- (1) You survive as Lefty.
- (2) You survive as Righty.
- (3) You survive as both Lefty and Righty.
- (4) You do not survive.

Parfit argues that none of (1)-(3) can be true, for reasons we have already discussed. The symmetry of the case seems to rule out (1) and (2), and (3) seems to be ruled out by the fact that Lefty  $\neq$  Righty.

- (1) You survive as Lefty.
- (2) You survive as Righty.
- (3) You survive as both Lefty and Righty.
- (4) You do not survive.

Parfit argues that none of (1)-(3) can be true, for reasons we have already discussed. The symmetry of the case seems to rule out (1) and (2), and (3) seems to be ruled out by the fact that Lefty  $\neq$  Righty.

Parfit is a psychological theorist. He thinks that the right answer is (4): you don't survive My Division.

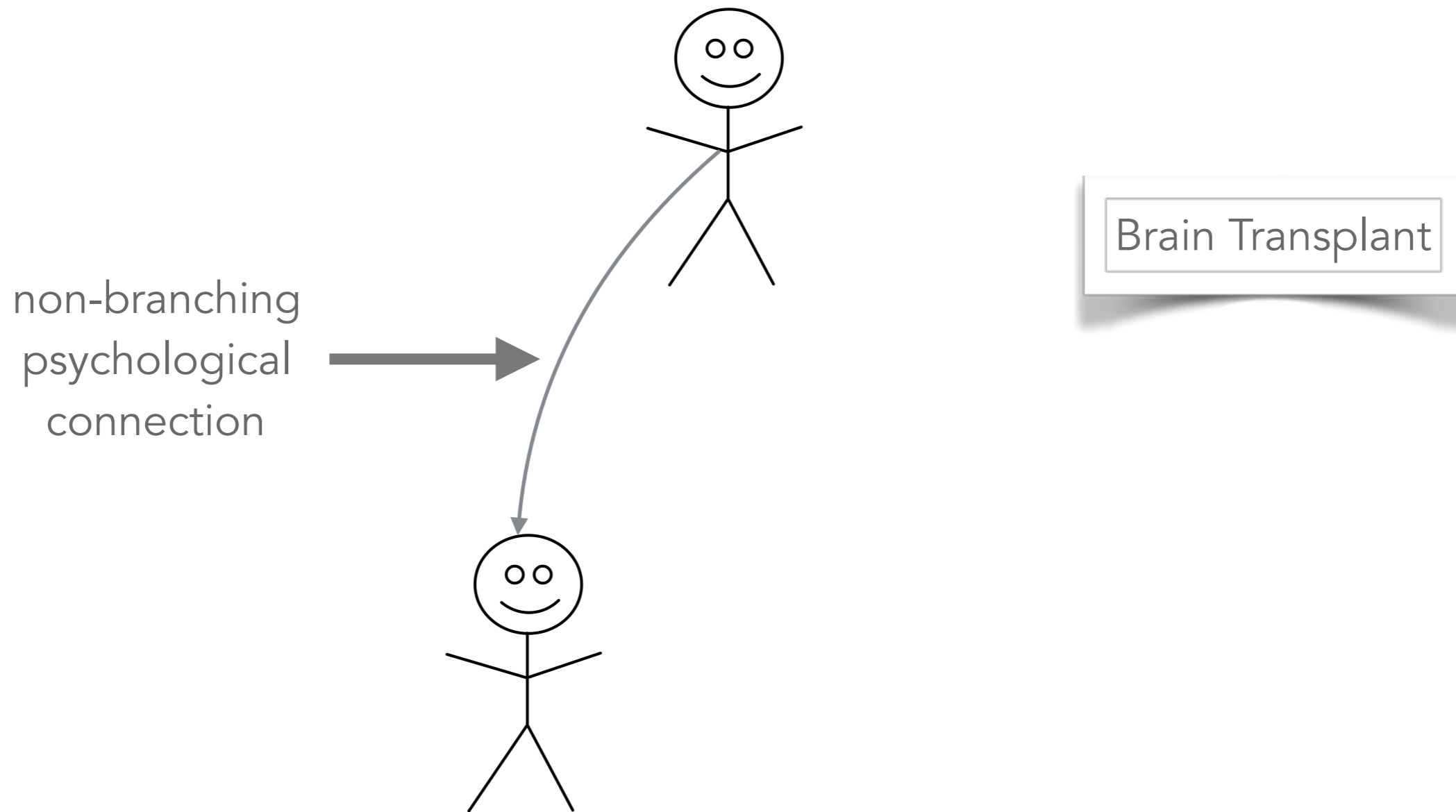
But how can he say this, since you stand in psychological relations to both Lefty and Righty?

Parfit says: the right kind of psychological connection is a **non-branching connection**. For you to survive at some later time, you must stand in a certain psychological connection to that thing and **to nothing else**.

That is why, he thinks, you can survive the Brain Transplant but not My Division.

Parfit says: the right kind of psychological connection is a **non-branching connection**. For you to survive at some later time, you must stand in a certain psychological connection to that thing and **to nothing else**.

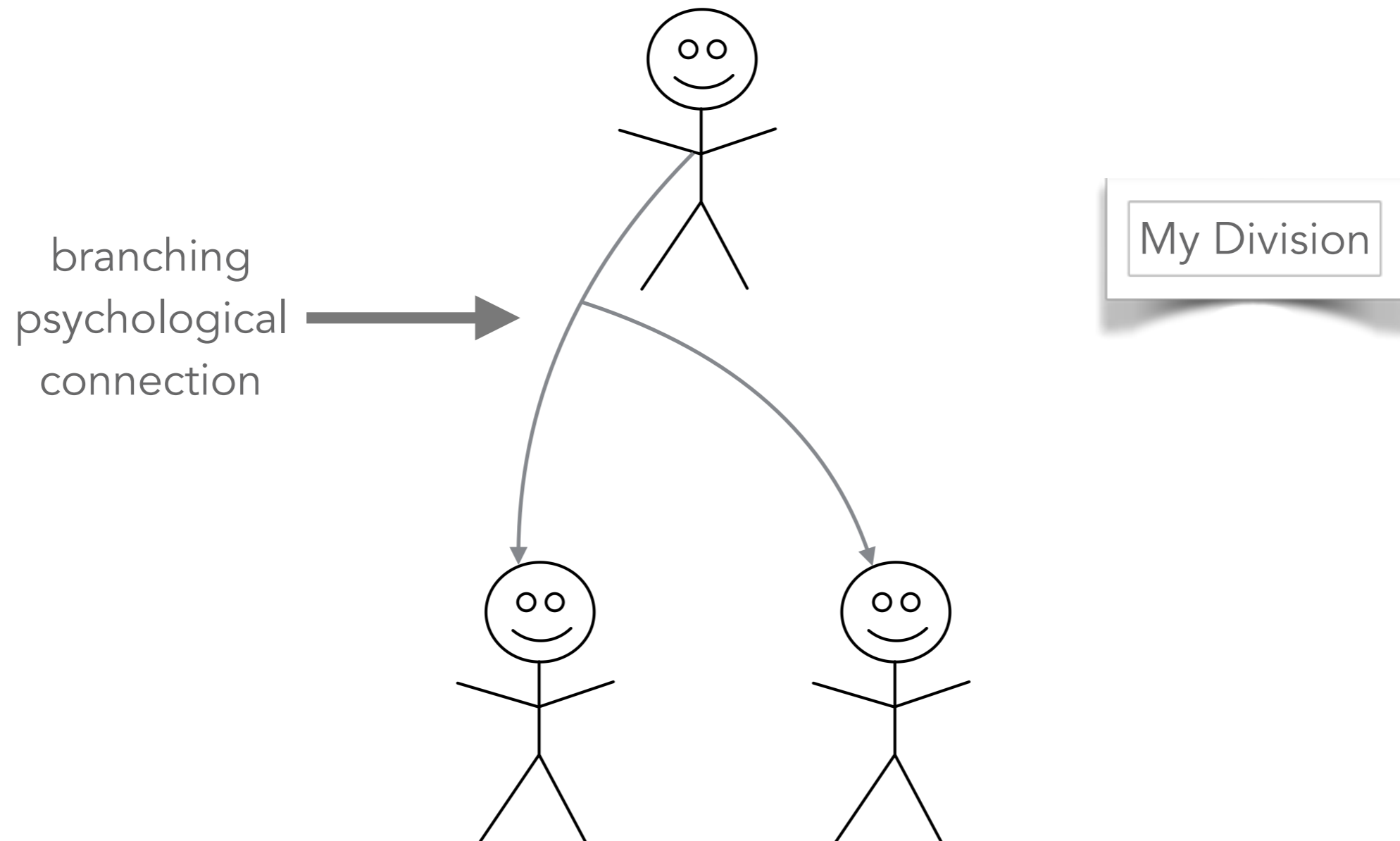
That is why, he thinks, you can survive the Brain Transplant but not My Division.





Parfit says: the right kind of psychological connection is a **non-branching connection**. For you to survive at some later time, you must stand in a certain psychological connection to that thing and **to nothing else**.

That is why, he thinks, you can survive the Brain Transplant but not My Division.



That is why, he thinks, you can survive the Brain Transplant but not My Division.

But, he thinks, things are just about as good for you in My Division as they are in Brain Transplant. True, you can survive the latter but not the former. But survival is not what matters; what matters is that something psychologically related to you survives, whether or not that thing is you.

Here is a test case:

Imagine that you suffer a devastating accident. The only way you can be saved is by individually taking the hemispheres of your brain and trying to attach them to bodies. Both surgeries are risky. You have the option of having this surgery tried for both of your hemispheres, or just for one. It seems obvious that you should have it tried for both, as this increases the chances of success. But if we think that personal identity is what matters, this would be a mistake. After all, you do not survive in the case where both surgeries are successful – so the probability of survival might be better if you do one surgery rather than two. So what do you choose?

Parfit thinks that reflection on these cases shows that we should care less about survival — and hence less about death — than we did before.

Thinking hard about these arguments removes the glass wall between me and others. And, as I have said, I care less about my death. This is merely the fact that, after a certain time, none of the experiences that will occur will be related, in certain ways, to my present experiences. Can this matter all that much?