

On Parfit's View That We Are Not Human Beings

Eric T. Olson, University of Sheffield

abstract

Derek Parfit claims that we are not human beings. Rather, each of us is the part of a human being that thinks in the strictest sense. This is said to solve a number of difficult metaphysical problems. I argue that the view has metaphysical problems of its own, and is inconsistent with any psychological-continuity account of personal identity over time, including Parfit's own.

1. The narrow criterion

Derek Parfit endorses a view of personal identity over time that he puts like this:

If some future person would be uniquely psychologically continuous with me as I am now, and this continuity would have its normal cause, enough of the same brain, this person would be me. If some future person would neither be uniquely psychologically continuous with me as I am now, nor have enough of the same brain, this person would *not* be me. In all other cases, there would be no answer to the question whether some future person would be me.¹

He calls this a 'narrow, brain-based psychological criterion'. It is 'psychological' by saying that someone can be me only if he is then psychologically continuous with me as I am now--or rather uniquely so, to rule out branching. 'Narrow' views say that our persistence requires psychological continuity with some physical qualification, for instance that one's basic mental capacities be 'continuously physically realized'.² 'Broad' psychological-continuity views, by contrast, say that anyone uniquely psychologically continuous with me, no matter how he got to be that way, is me. Suppose I underwent what Shoemaker calls a 'brain-state transfer': my brain is erased and its contents copied to your brain, resulting in someone psychologically like me but otherwise like you. This person would be me on a broad view but not on a narrow one, since my mental capacities would not be continuously realized: the machine simply destroys them and creates new capacities just like them in your brain.³ Parfit's version says that psychological

¹'We are Not Human Beings', *Philosophy* 87 (2012), 5-28

²Peter Unger, *Identity, Consciousness, and Value* (New York: Oxford University Press, 1990), 109

³My mental capacities don't exist in electronic form within the machinery, though the information represented in my mental contents does. For a good definition of psychological continuity, see Sydney Shoemaker, 'Personal Identity: A Materialist's Account', in Shoemaker and Richard Swinburne, *Personal Identity* (Oxford:

continuity has to be caused by the continued existence of my brain or a sizeable part of it, making it 'brain-based'.

In fact Parfit doesn't say that the brain-state-transfer recipient would not be me, but that there is no answer to the question of whether he would be.⁴ For someone *not* to be me, Parfit says, he must be neither uniquely psychologically continuous with me nor have enough of my brain. Thus, because the transfer recipient would be uniquely psychologically continuous with me but would have none of my brain, he would be neither me nor not me. The same goes for the being left behind with my original brain, now erased: he could not be me because he would not be psychologically continuous with me, and could not be distinct from me because he would have my brain.

The narrow criterion is an answer to the question of what it takes for us to persist through time. Parfit has recently argued from this and other considerations to an answer to another question: what are we? His proposal here that we are not human beings. Rather, each of us is the thinking part of the human being: something like the brain. It is this second view that I will be mainly concerned with. But the narrow criterion is interesting in itself, and I will start with it.

2. Peculiarities of the narrow criterion

Recall Shoemaker's brain-state transfer, where my brain is erased and its contents copied to another brain, resulting in someone uniquely psychologically continuous with me but with none of my brain. According to the narrow criterion, this person would be neither me nor not me. But if my brain states were copied to *two* beings, both recipients would definitely not be me. (None would be either uniquely psychologically continuous with me nor have enough of my brain.) Why should there be a definite fact about my survival in a double transfer but not in a single transfer? The reason is not that branching is always fatal: if each half of my brain were transplanted into another head, the narrow view implies that each resulting person would be neither me nor not me, since each would have enough of my brain. So I definitely don't survive in the double transfer, but there is no answer as to whether I survive in the single transfer or in the double brain transplant. Why this combination of verdicts?

Parfit's thinking may be that our persistence through time consists of two factors: unique psychological continuity and the persistence of enough of the same brain. If Blackwell, 1984), 89-91. The transfer story is from p. 108.

⁴When Parfit says that the question of whether someone would be me has no answer, I take him to mean no straight, yes-or-no answer. Rather, the person would be neither definitely me nor definitely not me, much as the answer to the question of whether it's raining may be that it's neither definitely raining nor definitely not raining. If this is right, then the question *has* got an answer--a unique correct answer, in fact--and Parfit's formulation is misleading. But I don't know what else he could mean.

both are present, you survive. If both are absent, you perish. If just one is present, you neither survive nor perish. The first factor is present in the single but not in the double transfer; the second is absent in both. So you perish in the double transfer but your fate is indeterminate in the single transfer. In the double transplant, the second factor--having enough of the same brain--is present, but not the first; so again it is indeterminate what happens to you. If this is right, we can put the narrow criterion more clearly like this (call it View 1):

A future person is me if he is uniquely psychologically continuous, then, with me as I am now, and he then has enough of my brain. A future person is not me if he is then neither uniquely psychologically continuous with me nor has enough of my brain. Otherwise there is no answer to the question of whether a future person would be me.

But whatever its attractions, View 1 appears to be inconsistent. Imagine a 'half-successful double-brain transplant'. Each of my cerebral hemispheres is transplanted into a different head. One recipient ends up psychologically continuous with me, but the second hemisphere is erased so that the one who gets it is not psychologically continuous with me. View 1 implies that the first recipient is definitely me, since he is uniquely psychologically continuous with me by having enough of my brain.⁵ And the two recipients are definitely distinct from each other. It follows that the second recipient, the one with the 'erased' hemisphere, is definitely not me. (If $x=y$ and $y\neq z$, then $x\neq z$.) Yet View 1 implies there is no answer to the question of whether the second recipient is me, since he has enough of my brain but is not psychologically continuous with me. (The narrow criterion as stated has the same implication.)

Parfit's formulation suggests a way of avoiding this problem. Maybe someone's having enough of my brain, by itself, does nothing to enable me to survive, but is relevant only if it supports psychological continuity. This gives the two factors, unique psychological continuity and having enough of the same brain, an unequal status: though survival requires both, the second makes a difference only when it accompanies and causes the first. That would give us something like this (View 2):

A future person is me if he is uniquely psychologically continuous, then, with me as I am now, and this is caused by his then having enough of my brain. A future person is not me if he is not then psychologically continuous with me.

⁵One cerebral hemisphere is clearly enough. No psychological-continuity theorist that I know of would deny that one could survive a hemispherectomy--a real though rare operation. And Parfit himself says that someone who was psychologically continuous with me by getting half my transplanted brain would be me if the other half were destroyed: see *Reasons and Persons* (Oxford: Clarendon Press, 1984), 254.

Otherwise there is no answer to the question of whether a future person would be me.

This avoids inconsistency by implying that someone who had enough of my brain but was not psychologically continuous with me would definitely not be me. But it isn't going to make anyone happy. Imagine a variant of the brain-state transfer that copies my states to another brain without erasing them from mine. View 2 implies that both resulting people would be neither me nor not me, since each would be psychologically continuous with me but not uniquely so. (View 1 and the original narrow criterion have the same implication.) Yet the procedure does me no more harm than an MRI scan. Surely I could survive it if I could survive anything.

We could block this implication by amending the first clause (View 3):

A future person is me if he is psychologically continuous, then, with me as I am now, this is caused by his then having enough of my brain, and no one else relates in these ways to me then. A future person is not me if he is not then psychologically continuous with me. Otherwise there is no answer to the question of whether a future person would be me.

On this view the person who had my brain at the end of the variant transfer would be me because he alone, after the transfer, would be both psychologically continuous with me and such that this was caused by his having enough of my brain. But this is again inconsistent. It implies that the recipient of the transfer (who would have none of my brain) would be neither me nor not me, since he would satisfy neither the first nor the second clause. Yet if I am definitely the one who has my brain after the transfer, and the one with my brain is definitely not the transfer recipient, then I am definitely not the recipient.

We could make View 3 consistent by changing its second clause (View 4):

A future person is me if he is psychologically continuous, then, with me as I am now, this is caused by his then having enough of my brain, and no one else relates in these ways to me then. A future person is not me if he is either not then psychologically continuous with me or hasn't got enough of my brain. Otherwise there is no answer to the question of whether a future person would be me.

In this case the variant-transfer recipient would satisfy the second clause, and thus would definitely not be me. View 4 has the important consequence that there can be indeterminacy only where two or more beings are at once psychologically continuous with me and have enough of my brain (or perhaps where psychological continuity or the like holds to an intermediate degree). Yet Parfit clearly wants to say that there is indeterminacy in a far wider range of cases--any where no amount

of reflection yields any strong conviction about who would be who.

What should Parfit say here? View 4 has no apparent advantage over the usual, industry-standard narrow psychological-continuity view:

A future person is me *iff* he is then psychologically continuous with me as I am now, this continuity is continuously physically realized, and at no future time do two people relate in this way to me as I am now.

This differs from View 4 by not requiring you to have the same brain to survive, and by implying that you definitely don't survive in the double brain-transplant case. More generally, it rules out any indeterminacy except in cases where the individual conditions hold to an intermediate degree. Whether that makes it more or less attractive than View 4 depends on why anyone might be drawn to Parfit's narrow criterion in the first place. That, unfortunately, is a question he never answers.

3. Why Parfit thinks we're not human beings

I turn now to Parfit's view that we are not human beings. Why does he say this? And why does he suppose instead that we are parts of human beings?

We *appear* to be human beings--that is, human organisms. I see a human being in the mirror. Isn't it *me* I see?

There is also a powerful argument for this view. Normal human beings can think. In fact they seem to have the same psychological properties that we have. But if they do, yet we are not human beings, it follows that there are two thinking beings wherever we thought there was just one. Human beings will also satisfy ordinary definitions of 'person', such as Locke's: the human being I see in the mirror is 'a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places'.⁶ So there are two different people thinking these thoughts. How could I ever know which one I am--the animal person or the nonanimal person? If I think I'm the nonanimal person, the animal person will be equally convinced, on the same grounds, that *he* is the nonanimal. No possible evidence, it seems, could favour one alternative over the other. Even if I were not a human being, it's hard to see how I could ever know it. The obvious solution to this problem of 'too many thinkers' is to say that we are human beings.

Most of those who deny that we are human beings try to avoid the problem by saying that human beings are not psychologically just like we are: despite appearances, they don't think, or satisfy Lockean definitions of personhood. But that is hard to defend. What could prevent a normal human being from using its brain to think? If human beings don't think, that can only be because it is metaphysically impossible for a living organism to have any mental property at all.

⁶John Locke, *An Essay Concerning Human Understanding*, ed. P. Nidditch (Oxford: Clarendon Press, 1975), 335

The property of being alive--alive in the sense that organisms are alive--must be incompatible with the property of having beliefs or being conscious. What appears to be a living, conscious being would have to be a compound of a living but unconscious being and a conscious but nonliving being. This is a sort of substance dualism--not Descartes' dualism of mind and matter, but a monstrous dualism of mind and life.

Parfit denies that we are human beings for three reasons. The first is familiar: our being human beings--'animalism'--has counterintuitive consequences about our persistence--consequences inconsistent not just with the narrow criterion but with any psychological-continuity account of personal identity over time.⁷ Suppose your brain were put into my head. The resulting person would remember your life, and not mine. He would have your beliefs, preferences, plans, and other mental properties, for the most part at least. Who would he be? Animalism implies that he would be me with a new brain. The operation simply moves an organ from one animal to another, like a liver transplant.⁸ There are two human beings in the story; one of them loses its brain and becomes an empty-headed vegetable or corpse, and the other has its brain removed and replaced with yours. According to animalism, you are the donor organism and I am the recipient. You get an empty head; I get your brain. I would wake up mistakenly convinced that I was you.

Most people find it more plausible to suppose that the one who got your transplanted brain would be you. But no human being would go with its transplanted brain. It would follow that even though you are never actually going to have a brain transplant, you have a property that no human being has, that of *possibly* going with your transplanted brain. And in that case you are not a human being. This is the transplant problem.

Parfit's second objection to our being human beings is subtly different.⁹ Suppose your brain is removed from your head as before, only this time it's kept alive in a vat rather than transplanted. It may be possible for your brain in this condition to think more or less normally. (This is contentious, though many philosophical thought experiments assume it. I won't dispute it here.) That would make it a person. This 'remnant person' would not be a human being--that is, an organism. It would be alive only in the way that a kidney awaiting transplant is

⁷'We are Not Human Beings', 9-11

⁸If an entire detached brain *would* be an organism (see Peter van Inwagen, *Material Beings* (Ithaca: Cornell University Press, 1990), 169-181, let only the cerebrum be transplanted. This applies also to the remnant-person problem described below.

⁹Mark Johnston, "'Human Beings' Revisited: My Body is Not an Animal', in Dean Zimmerman, ed., *Oxford Studies in Metaphysics* 3 (Oxford: Oxford University Press, 2007), 33-74; Parfit, 'We Are Not Human Beings', 11-12; Eric Olson, 'The Remnant-Person Problem', in Stephan Blatti and Paul Snowdon, eds., *Essays on Animalism* (Oxford: Oxford University Press, forthcoming).

alive--in that its individual cells are kept alive. Nor was it previously a human being: there was only one human being there before your brain was removed, and the operation merely gave it an empty head.

That some people are not organisms is perfectly compatible with *our* being organisms. It could be that we are human beings and God is an immaterial person. But if some *human* people were not organisms, we ought to wonder why there isn't such a person associated with every human being. If an inorganic human person could live in a vat, why not in a head? And if there were such a person in my head, he would either be a second person in addition to me, which would be absurd, or he would be me, in which case I should not be a human being.

Animalists will also find it hard to say where the remnant person could have come from. If she existed (and was a person) before the operation, they will have to say that there were two people within your skin: the human being who got an empty head and the remnant person who went into the vat. Presumably there would again be two people wherever we thought there was just one. If the remnant person did not exist before the operation, on the other hand, then the operation must have brought her into being. But how can you create a new person just by cutting away sustaining tissues?

It's hard to find a plausible account of the nature of remnant people and of what happens to them in this case that is compatible with animalism, or with any other view according to which we are the size of human beings. This is the *remnant-person problem*.

Finally, Parfit says, animalism faces its own problem of too many thinkers: the *thinking-parts problem*.¹⁰ Isn't your *brain* conscious and intelligent? It would be (we have supposed) if it were kept alive in a vat. But that could hardly *make* your brain conscious and intelligent: it couldn't be that the surrounding tissues prevent the brain in your head from thinking or being conscious, and removing it from its natural habitat would suddenly give it those capacities. If so, then your brain must be conscious and intelligent even now. It must be a person by any ordinary definition. Yet if you are a human being, your brain isn't you. Animalism appears to imply that you have a smaller person within you. How, then, could you ever know whether you are the animal person or the brain person? This looks no less worrying than the original thinking-animal problem.

4. The embodied-part view

Parfit says there is 'an obvious solution' to all four problems: the thinking-animal problem, the transplant problem, the remnant-person problem, and the thinking-parts problem. It is that we are not human beings, but parts of them. Specifically, each of us is the part that thinks our thoughts in the strictest sense: we are 'the conscious, thinking, and controlling parts of human beings'.¹¹ Parfit

¹⁰'We Are Not Human Beings', 13f.; see also Eric Olson, *What Are We?* (New York: Oxford University Press, 2007), 215-219.

assumes that an organism thinks by having a smaller part that thinks--the brain or some part of the brain--much as a locomotive is powerful by having a powerful engine as a part. (I take 'thinking' to include all mental activities and properties.) The organism has its mental properties derivatively. But not every part of a thing can think only derivatively: it can't be that an organism thinks by having a smaller part that thinks, and that smaller part thinks by having an even smaller thinking part, and so on forever. At some point there must be something that thinks in its own right. That thing is the person.

So the proposal makes three claims: (1) for every thinking organism, there is just one nonderivative thinker, which (2) is a part of the organism and (3) is the person. Parfit calls this the *embodied-part view*. (The name, I presume, alludes to the fact that in normal cases we are 'attached' parts of human beings rather than detached or 'disembodied' ones kept in a vat or the like. 'Thinking-part view' would be a more perspicuous name, as it would tell us what part of the human being each of us is. But I will adopt Parfit's terminology in order to avoid confusion with the thinking-parts problem.)

This would solve the thinking-animal problem by implying that human organisms don't really think. (I will discuss why not in the next section.) In a way, perhaps, the animal is a thinker in addition to me. But it thinks only insofar I am a part of it, whereas I think independently of my parthood relations to other thinkers. There is only one true, nonderivative thinker of my thoughts. I can know that I am not the human being because I know that I am the true thinker of my thoughts--arguably it belongs to the content of first-person thoughts that they refer to the being who thinks them nonderivatively.

It would solve the transplant problem by implying that you would go with your transplanted brain. You are the brain (or some part of it) all along. The operation does not change your size, but literally moves you from one head to another, like repotting a plant.

It would solve the remnant-person problem by saying that the remnant person is you. Again, the operation simply moves you from your head to the vat.

And it would solve the thinking-parts problem by saying that there is only one part of a human being that thinks nonderivatively: me. I have no thinking parts.

The proposal has a further advantage over animalism that Parfit doesn't mention. Some say that in special cases of conjoined twinning, a single human being has two brains that function as independently as mine and yours.¹² In such cases, they argue, there are two different people. So at least one of those people, and probably both, must be something other than a human being. But if a two-headed human being would 'contain' two people who were not animals, we should expect each ordinary human being to contain one nonanimal person--someone of

¹¹'We Are Not Human Beings', 14

¹²Tim Campbell and Jeff McMahan, 'Animalism and the Varieties of Conjoined Twinning', *Theoretical Medicine and Bioethics* 31 (2010), 285-301

the same metaphysical nature as the 'twin people'. So animalism itself implies that there are two people wherever we thought there was just one, an animal person and a nonanimal person. The embodied-part view can solve the problem by saying that there really would be two people in the twinning cases: they would be the two brains.¹³

Solving these five hard problems is a great merit, and the embodied-part view deserves a hearing. Its implication that we are just a few inches tall, or that few of us have ever really seen ourselves or anyone else, would be a small price to pay. I think the view faces far more serious metaphysical objections. It also looks incompatible with anything like Parfit's narrow criterion. Alterations to make it compatible create further problems, as well as diminishing its problem-solving virtues.

5. Thinking-subject minimalism

Which part of the organism is the nonderivative thinker, the 'thinking part'? Perhaps we can assume that it would have to include at least some of the brain. It couldn't be a hand or a finger. But that is consistent with its being the animal's upper half, or the head, or the entire brain, or this or that part of the brain. Which is it? For that matter, why not the entire organism? Why must an organism have a smaller part that thinks nonderivatively? Why, in other words, should it be impossible for an organism think in its own right?

The embodied-part view implies that this question must have an answer. If there were no saying what part of the organism was the thinking part, there would be no saying what we are--whether heads or brains or even whole organisms. That would be incompatible with the claim that we are not human beings.

Parfit says little about this. He suggests that the thinking part is no larger than a brain, but never says why. I suppose the answer must be something like this: a true thinker has to be made up of all and only the objects directly involved in its thinking. Human beings cannot think because they have superfluous parts--feet, for instance--that have no direct involvement in their mental activities (or rather, those going on within them). Having feet as parts is metaphysically incompatible with thinking. So my feet cannot be parts of me. Call this principle *thinking-subject minimalism*.¹⁴ Without it or something similar, it would be entirely arbitrary to say that I am a brain rather than some other part of an animal, or indeed a whole

¹³Animalists have their own responses to these problems. On the transplant objection, see Olson, *The Human Animal* (New York: Oxford University Press, 1997), 42-69; on the remnant-person problem, see Olson, 'The Remnant-Person Problem'; on the thinking-parts problem, see van Inwagen, *Material Beings*, 81-97, Olson, *What Are We?*, 215-219; on the twinning problem, see van Inwagen, *Material Beings*, 188-212, Olson 'The Metaphysical Implications of Conjoined Twinning', *Southern Journal of Philosophy* 52 (2014), 24-40.

¹⁴Olson, *What Are We?*, 87-90

animal.

I think minimalism is trouble. As far as I know, the only argument for it is that it would help solve the five problems (if that counts). It certainly gets no support from Parfit's narrow criterion, or from his view that we are essentially thinking beings. In fact on closer examination it looks incompatible with them.

Consider first how it could be generalized. Minimalism can hardly be just a principle about thinking. If an animal cannot think because it has parts not directly involved in its thinking, something analogous ought to hold for other activities. Because an animal has parts not directly involved in its walking, for instance, it must walk only in a derivative sense. There must to be a unique part of the animal that walks strictly speaking: its walking part. It will presumably be different from the part that talks, the part that eats, and the part that sleeps, since the things directly involved in an animal's walking will not be those directly involved in its talking, eating, or sleeping. And these other parts will be different from each other. What we think of as a single being that does many things must really be an assemblage of many beings that each do only one thing. I, the thinking part, never do anything but think. It would be impossible for me to walk or talk or eat, except in the loose and derivative sense of being a part of an animal that has walking, talking, and eating parts. I only give orders; other beings, which are not even parts of me, carry them out. In this respect minimalism resembles the Cartesian view that I am an immaterial substance.

If that isn't worrying enough, it will be hard to say what parts of an organism are directly involved in thinking or any other activity, as opposed to only indirectly involved or not involved at all. Since I couldn't think unless my brain had a supply of oxygenated blood, my heart and lungs are involved in my thinking. But if the thinking part of the animal is no bigger than the brain, they're not directly involved. (Neither, presumably, are those parts of the brain that deliver the blood.) Why not? What *is* it for something to be directly involved in someone's mental activities?

Walking is easier to visualize. Which part of an animal is the walking part? When I swing my arms as I walk, are they *directly* involved in my walking, as my feet are, or only indirectly involved? What could settle this question? And even my feet have parts that make no evident contribution to my walking--toenails, for instance. At most only certain parts of my feet seem directly involved. But which ones? Suppose I have excess water in my feet owing to poor circulation, hindering my walking. In that case it seems that not all the water molecules in my feet would be directly involved in my walking, though some must be. Yet it's not as if some of these molecules are the excess ones. What hinders my walking is not the presence of particular molecules in addition those that 'belong there', but simply that there is too much water.

I doubt whether there is any principled way of saying which molecules are directly involved in my walking, which are only indirectly involved, and which are not involved at all. The point is not merely epistemic. Not even complete

knowledge of the microphysical details would tell us which are which. Nor is it merely that some molecules are neither definitely directly involved in my walking nor definitely not directly involved. If we are material things of any sort, we're bound to have fuzzy edges. The problem is that there is no principled way of drawing even a vague boundary around the walking part of me. And thinking is unlikely to differ in this respect from walking. Minimalism appears to imply that there is no saying which things are parts of me, and thus no saying which thing I am.

But there is trouble for minimalism even if there really is an absolute distinction between the things directly involved in thinking and those not directly involved. If some things are directly involved in my thinking generally, some will be directly involved in specific mental activities. Some neurones will be directly involved when I imagine people's faces; others--different ones--will be directly involved when I try to remember their names. And any reason to suppose that a true thinker must be composed of just those things directly involved in its thinking looks like an even better reason to suppose that a true imaginer or rememberer must be composed of just those objects directly involved in its imagining or remembering, respectively. The thing that imagines faces will be either too big to remember names, by having parts not directly involved in that remembering, or too small, by not including such parts--or both, of course. More generally, if it makes sense to speak of a neurone's being directly involved in thinking, then every mental activity will involve different neurones, and must therefore be performed by a different part of the organism. These parts may overlap. There may even be a part they all share (the reticular activating system, perhaps). But they are nevertheless different objects.

It follows that a human being hasn't got one nonderivative 'general' thinker within it, but rather a crowd of narrow specialists: one that only imagines, another that only remembers, a third that does nothing but think about metaphysics, and so on. What we take to be a person able to perform all sorts of mental operations is really many beings, each able to perform only one. If, as Locke said, a person is by definition both intelligent and self-conscious, then there are no people, and we do not exist. That is incompatible not only with Parfit's embodied-part view, but with any familiar account of what we are.

6. The embodied-part view and identity over time

But set aside worries about thinking-subject minimalism. Suppose each human being really has got a unique thinking part. For the sake of simplicity, let it be the brain, though the point is the same whatever part of the organism it might be.

Physiologists tell us that my brain came into being early in gestation, long before it could support any mental activity. If I am my brain, as the embodied-part view appears to imply, then I must have persisted then without any psychological continuity. That is incompatible with any view on which psychological continuity is necessary for me to persist, including Parfit's narrow criterion.

Or imagine that when I die my brain is fixed in formaldehyde and kept in a jar. It looks as if my brain would still exist then (and the same would go for any other part of the organism that I might be). That is, the brain in the jar is the brain that was once in my head. But if my brain could be pickled, and I am my brain, then I could be pickled. Again, I could persist without being psychologically continuous with anything.

Someone might argue that the pickled brain was never in my head, but is a brand-new object created by the fixing process, and that the embryonic, precognitive brain does not survive to maturity, but ceases to exist when it begins to produce thought and is replaced by a new brain. But that would make our beliefs about the persistence of brains badly unreliable, casting doubt on the assumption that my brain could continue existing when transplanted or kept alive in a vat. If *we* are brains, that would undermine all claims about our persistence through time.

Now the narrow criterion as Parfit states it speaks only of what it takes for a future *person* to be me. It says nothing about future *nonpeople*, never mind past ones. It tells us what it takes for someone to persist *as a person*, but not what it takes for someone to persist in general. If my pickled brain would not count as a person (it certainly wouldn't then satisfy anything like Locke's definition), that might seem to make the narrow criterion compatible with my existing in the jar.

I think Parfit stated his view as he did because he assumed, like other psychological-continuity theorists, that a person is a person essentially, and none of us could exist as a nonperson. On that assumption it would make no difference if we replaced 'future person' in his account with 'future being': the two formulations would be logically equivalent. If Parfit really did mean to tell us only what it takes for someone to persist as a person, his narrow criterion would be radically incomplete. It would tell us nothing about what would happen to us in a wide range of cases. It would allow me to become a lifeless corpse, or even a laurel tree or a pillar of salt. It would be compatible with any claim at all about what sort of thing I could come to be, as long as I was not then a person. Nor would it tell us when we came into being: whether at the earliest point when there was the right sort of psychological continuity, at fertilization, or even earlier. I doubt whether that was Parfit's intention. It certainly isn't that of other psychological-continuity theorists.

7. Functioning brains

Everything Parfit has ever said about personal identity rules out my existing in formaldehyde. If my brain could exist in formaldehyde, then I am not my brain. What am I, then? Parfit doesn't say. He makes only the suggestive remark that people stand to their brains as animals stand to their bodies.¹⁵ The thought seems to be that the atoms making up an organism--a fish, say--also compose a second

¹⁵'We Are Not Human Beings', 15. He does contrast our being brains with our being 'embodied minds' or 'embodied persons', saying that he prefers the last (15-17). But what these proposals amount to and how they differ is never explained.

material thing that is not a fish, but rather the fish's body. This object is empirically indistinguishable from the fish for as long as the two coexist. They have the same shape, size, mass, chemical composition, anatomical and cellular structure, and so on. Yet they are different. One difference is that when the fish dies, it ceases to exist, while its body carries on in a nonliving state.

So an organism shares its matter with another object, its body, having different persistence conditions. And so does a human brain (supposing that that is the right organ): it shares its matter with the conscious, thinking, controlling part of a human being. When the brain loses the capacity to support thought, it simply ceases to function and carries on in a nonfunctioning state (like the fish's body); but the other thing (like the fish) vanishes. We might call this other thing the *functioning brain*. Since we are functioning brains and not brains, we cannot not exist in formaldehyde. Nor did we exist in an unconscious embryonic state.

I am sceptical about Parfit's metaphysics of organisms and their bodies.¹⁶ I am even more sceptical about functioning brains. Why suppose that a brain shares its matter with an object physically identical to it but with different modal properties? Consider the claim that there is such a thing as my *waking brain*: a thing just like my brain except that it can exist only as long as it is awake. When I fall asleep, my waking brain instantly vanishes (replaced, perhaps, by a 'sleeping brain' that is essentially asleep). It returns to being when I wake up again. I can see no reason to believe in the existence of functioning brains that is not equally a reason to believe in the existence of waking brains. And if there are waking brains, there are probably also things just like brains except that they are essentially thinking about philosophy, essentially sober, or essentially north of the equator, and likewise for every combination of such properties. Where is this going to end? Is there, for every property my brain currently has, a thing just like it but having that property essentially?

That would be an extraordinary metaphysical claim, and we ought to accept it only on the basis of an extraordinary argument. Parfit gives no argument at all. In any event, the claim raises a problem of too many thinkers infinitely worse than those the embodied-part view was supposed to avoid. If there is any interesting connection between the mental and the physical, we should expect physically identical objects in physically identical surroundings to have the same mental properties.¹⁷ So my waking brain, my sober brain, and all the other objects of their ilk ought to be psychologically identical to my functioning brain. If my functioning brain is a person in Locke's sense, then my waking brain and my sober brain will

¹⁶I discuss some better alternatives in 'The Person and the Corpse', in Ben Bradley *et al.*, eds., *The Oxford Handbook of Philosophy of Death* (New York: Oxford University Press, 2013), 80-96.

¹⁷Or at least the same intrinsic mental properties: my doppelgänger in another galaxy may think about a city physically identical to Vienna, but he cannot think about Vienna.

be too. There will be a vast number of people--probably an uncountable infinity--wherever we thought there was just one. How could I know which person *I* was? For all I could ever find out, it seems, I might be one who can exist only as long as I continue writing this sentence. In that case I could never know anything about what it takes for me to persist through time.

Maybe there are functioning brains but no essentially waking brains and the like. But even if this mysterious claim were true, there would still be my brain itself (or some part of it): a thing physically and mentally identical to me, yet not me. It would be a second person within my skin. How could I know whether I was the brain person who could be pickled or the functioning-brain person who couldn't be? This is a version of the thinking-parts problem that Parfit thought he had solved. For that matter, our brains would be human people able to persist without psychological continuity, contradicting Parfit's narrow criterion.

Parfit will need to say that the functioning brain is the only thinking part of the organism, and that the brain itself thinks only in the loose and derivative sense of relating in some way to the functioning brain. Each normal human being contains only one person. But how could there be two physically identical objects, in identical surroundings, only one of which can think? The embodied-part view says that human beings cannot think because they have parts not directly involved in thinking. But this does nothing to explain why my brain can't think: it has exactly the same parts as my functioning brain, which by hypothesis are just those parts of the organism directly involved in my thinking.

So Parfit cannot solve the new thinking-parts problem in the way that he would solve the thinking-animal problem, by saying that my brain doesn't think because it has parts not directly involved in thinking. Nor can he solve it in the way that he would solve the original thinking-parts problem, by saying that I *am* my brain, as that is incompatible with the narrow criterion.

8. Alternatives

The best way of salvaging Parfit's views may be to say that my functioning brain is a *temporal* part of my brain (or of the relevant spatial part of my brain). The assumption here is that for every matter-filled region of space-time, there is a concrete material thing--just one--occupying precisely that region. This implies that every part of my brain's career is the career of a temporal part of my brain. This part, for as long as it exists, is physically just like my brain. My functioning brain would then be the mereological sum of those temporal parts of my brain that function in the appropriate sense--or, better, the largest sum of its temporal parts, each of which is psychologically continuous with every other. It would have come into being several months after my brain did, when, in the course of my foetal or infant development, the right sort of psychological continuity first began. If my brain is fixed in formaldehyde, its preserved temporal parts will not be parts of my functioning brain either. And Parfit might take thinking-subject minimalism to apply

to temporal as well as spatial parts, so that every temporal part of a thinker must be directly involved in its thinking. That would enable him to say that my brain cannot think for much the same reason that a human being cannot: because it has parts not directly involved in thinking (its foetal and pickled temporal parts). This would solve the thinking-parts problem--though it would retain the troubling commitment to thinking-subject minimalism. It is the view defended in Hud Hudson's fine book *A Materialist Metaphysics of the Human Person*.¹⁸

Parfit often argues in a way that seems to presuppose an ontology of temporal parts. For instance, he says that if there are such things as people, there must also be such things as 'day-people', who are just like people except that they must have an uninterrupted stream of consciousness.¹⁹ There is a day-person now writing these words, who will cease to exist when I next fall asleep. If you and I were composed of arbitrary temporal parts, some of them really would be day-people, and their existence would follow from our own. Otherwise their existence looks eminently doubtful. But over the course of his career Parfit has declined every opportunity to endorse an ontology of temporal parts.

Alternatively, he could give up the narrow criterion and say that we are brains in the ordinary sense. That would avoid both the thinking-parts problem and the mysterious ontology of functioning brains. It would imply, of course, that our persistence consists in some sort of brute-physical continuity. I really could be pickled. I would go with my brain if it were transplanted, but not because that organ secures psychological continuity. If you don't like that, Parfit could point out that *all* views of personal identity over time have unattractive consequences, and remind us that numerical identity is unimportant: whether a past or future being is me or someone else is not itself something that I have any reason to care about. What matters practically is some sort of psychological continuity.

Yet another possibility would be to say that we are immaterial, and solve the too-many-thinkers problems by saying that no material thing could ever think or be conscious. That would mean giving up both the narrow criterion and the embodied-part view, though it would have the considerable advantage of not requiring thinking-subject minimalism or any controversial claims about the ontology of material things.

Or he could concede that we are human beings after all. I cannot see that he has any better options than these.²⁰

¹⁸Ithaca: Cornell University Press, 2001

¹⁹*Reasons and Persons*, 292

²⁰For comments on ancestors of this paper I thank Radim Belohrad, Stephan Blatti, Galen Strawson, and Alex Moran. I am also grateful to Derek Parfit for discussions of these and related matters.