DE GRUYTER OPEN

Perichoresis
Volume 15. Issue 2 (2017): 3–24
DOI: 10.1515/perc-2017-0007

THE SOUL AND PERSONAL IDENTITY. DEREK PARFIT'S ARGUMENTS IN THE SUBSTANCE DUALIST PERSPECTIVE

DMYTRO SEPETYI*

Zaporizhzhya National University

ABSTRACT. This paper re-evaluates Derek Parfit's attack on the commonly held view that personal identity is necessarily determinate and that it is what matters. In the first part we first argue against the Humean view of personal identity; secondly, we classify the remaining alternatives into three kinds: the body theory and the brain theory, the quasi-Humean theory, and the soul theory, and thirdly we deploy Parfit's arguments and related considerations to the point that none of the materialistic alternatives is consistent with the commonly held view. This leaves us with the alternative: either we accept the radical and highly implausible materialistic view Parfit calls 'Reductionism', or we accept the view that we are nonphysical indivisible entities—Cartesian egos, or souls. The second part of the paper discusses Parfit's objections against the Cartesian view: that there is no reason to believe in the existence of such nonphysical entities; that if such entities exist, there is no evidence that they are enduring (to span a human life); that even if they exist and are enduring, they are irrelevant for the psychological profile and temporal continuity of a person; that experiments with 'brain-splitted' patients provide strong evidence against the Cartesian view. We argue that these objections are in part mistaken, and that the remaining (sound) part is not strong enough to make the Cartesian view less plausible than Reductionism.

KEY WORDS: personal identity, soul, materialism, dualism, substance

Introduction

I am looking at this computer screen and thinking of the visual experiences I have. It is the same mental subject, me, who has these experiences and who thinks these thoughts. I will call this kind of sameness *personal identity at a time*, or *synchronic personal identity*. I can recollect some of my experiences and thoughts that I had several years ago, and I take for granted that I had lots of other experiences ant thoughts which I cannot recollect now. It is the same mental subject, me, who thinks these thoughts now and who had

* DMYTRO SEPETYI (PhD 2009, Zaporizhzhya National University, Ukraine) is Assistant Professor at the Department of Social Sciences, Zaporizhzhya State Medical University, Ukraine. Email: dmitry.sepety@gmail.com.

those experiences ant thoughts then. We will call this kind of sameness *personal identity over time*, or *diachronic personal identity*.

With all of us (mentally sound human beings), our everyday thinking is permeated with the presumptions of these identities, which are taken for granted. Moreover, these presumptions are implicit in the main bulk of what we take as worth bothering about, as things that matter. If we bother about life and death, suffering and pleasure, we are usually concerned with the continuation and the subjective quality of the existence of ourselves or others as those who suffer and/or enjoy their continued existence, that is, as enduring mental subjects. This holds equally for 'egotistic' concerns of a person about his or her own life, death, suffering, and pleasure and for altruistic concerns of a person about other human or animal beings' life, death, suffering, and pleasure *for their own sake*.

However, there is a problem about those supposedly enduring mental subjects that we take ourselves to be: the belief that there are such subjects and that we are such subjects sits badly with the materialistic view of the world and of the human beings, which seems to be amply supported by science and is accepted by most of contemporary academic philosophers. To see just how badly these views sit together, it is very helpful to read Derek Parfit's book *Reasons and Persons* (part three, Personal Identity). Stephen Law's popular rehash of these issues (based for the largest part on Parfit's discussion) in his bestseller *Philosophy Files* (story 'Where am I'), is also helpful.

The most relevant parts of Parfit's discussion were titled 'What we believe ourselves to be' and 'How we are not what we believe'. For the largest part, we agree with the former and disagree with the latter. In what follows, we (1) restate and critically survey the alternative theories of personal identity and (2) argue that—*pace* Parfit—there are good reasons to think that we are what we believe, that is, absolutely self-identical (in Parfit's terms, determinate) mental subjects, and that this means that we are non-physical indivisible substances, or selves, or souls.

Making Alternatives Clear

We propose to begin with several points that seem unproblematic.

(1) We talk of our selves, using first-personal pronouns, and when we do this, we mean either ourselves as mental subjects (ones who experience, think, will, etc.), or our bodies.

(2) These two meanings are distinct: when talking or thinking of ourselves as mental subjects, we need not necessarily think of our bodies, and when talking or thinking of our bodies, we need not necessarily think of our selves as mental subjects; we can conceive of the possibility of ourselves be-

ing un-embodied, and of our bodies running as automata (phenomenal zombies) without ourselves, as mental subjects, involved, etc. For a while, we leave it open whether the things so conceived—mental subjects and human bodies—are in fact distinct or identical. Our present claim, which we take to be unproblematic, is just that the concepts, their meanings, are distinct.

3. The central meaning of our talks of ourselves is that of mental subjects. The talk of our bodies as ourselves is derivative: we talk, in some contexts, of our bodies as ourselves only insofar as we see them as sources of our experiences and obedient servant of our will.

If these points are granted, it seems that we should proceed on the assumption that our personal identities are identities of those mental subjects that we are. The remaining substantial issues about personal identity are: 'What those subjects are?' ('What is their nature?') and 'What it takes for a mental subject to endure, to remain itself for a span of time?' ('What is the criterion of a mental subject that exists at time t being the same with a mental subject that existed at some previous time t_0?')

However, we can proceed on that assumption only *if there are mental subjects at all, if there is something in the world that corresponds to our concept of a mental subject, or self*. This is challenged by a highly influential view that was famously advanced by David Hume (Hume 1951: 251-263). On this view, there are no mental subjects, or selves, in the sense we are used to—no mental subjects as distinct from their experiences (or other mental states), as 'havers' of those experiences (mental states). Of all there really is, the closest approximation to our concept of a self is a temporally continuous bundle, or stream of mental states. So, we should think of ourselves as nothing but such bundles, or streams.

The Humean view was, and still is, highly influential. One historically important example of the influence of this view was George Lichtenberg's objection to Descartes' *Cogito*: Descartes should not have claimed 'I think, therefore I am'; he should claim instead 'It is thought: thinking is going on' (Lichtenberg 1971: 412). Later, this objection was repeated by Bertrand Russell (Russell 1945: 567), whose impact on the development of 20th century analytical philosophy can hardly be overestimated. The mainstream of the contemporary philosophy of mind also makes impression of being strongly influenced by the Humean view. Although there are not many philosophers who *explicitly* subscribe to the view that experiences (mental states) can exist without experiencers (mental subjects), the view seems to be implicit in much, if not most, of the discussion in the field, in that it is centred around experiences but pays little or no attention to experiencers, and many eminent philosophers often talk of experiences and other mental

states as if they are things that exist (or are capable to exist) on their own, as no one's experiences or mental states.

Despite such authorities and the mainstream, we think that the Humean view is grotesquely untenable. The idea that our selves are nothing but streams of experiences (mental states) implies that experiences (mental states) are distinct entities capable of existing on their own, without there being any mental subjects (selves). For a mental subject (self) to appear, a number of such no-one's experiences (mental states) should get arranged into a continuous stream. However, the idea of no-one's experiences (mental states) is just as good as the idea of a round square. Being someone's is inalienable from the very concepts of experiences, thoughts, and other mental states. If we subtract this 'someoneness' from the meanings of such mental concepts, nothing is left. The talk of experiences without an experiencer, thoughts without a thinker, etc. is just unintelligible abracadabra. We suggest that this is an excellent reason to reject the Humean view of personal identity and take the existence of mental subjects for granted.

Now we can turn to the further substantial issues, of what those mental subjects are and what their diachronic identity consists of. We think that all the alternative views on these issues can be conveniently divided into three major kinds:

(1) *The body theory of personal identity and the brain theory of personal identity.* A mental subject is (identical with) a human body or part thereof (supposedly, the brain), and its diachronic identity consists in the moment-to-moment physical continuity of the body (brain).

(2) *The quasi-Humean theory of personal identity.* A mental subject at a time is (identical with) a human body or part thereof (supposedly, the brain), and its diachronic identity consists in the moment-to-moment psychological continuity of instantaneous bodies (brains).

(3) *The soul theory of personal identity.* A mental subject is distinct from a human body, and any part thereof; it is a non-physical entity (usually called 'soul'), and its diachronic identity consists in that entity's continued existence.

The first kind of theory is the most obvious option for a materialist. The third theory is the view of substance dualism, or the Cartesian view.

Probably, the second theory and its credentials need explanation. The theory is materialistic with a Humean turn. Why a materialist can prefer such a theory rather than be satisfied with a more straightforward materialistic option—the body theory or the brain theory? To see the reason, the comparison of two varieties of theories of the first kind is helpful.

There is the body theory (I am my body) and the brain theory (I am my brain) of personal identity. Which one of the two is preferable and why?

In reality, the human body continuity and the human brain continuity were always going together. So far, there were no cases in which they diverge (such as replacing a brain in a body with another brain, or creating a new body for an old brain). However, such cases are possible in principle, and perhaps will be practicable in some future. The simplest imaginable case is that of swapping brains between two human bodies.

Imagine a surgical operation in which Bill's brain is transplanted into John's body, and John's brain is transplanted into Bill's body. If such a swapping were made, would personal identity go with the brain, or with the body? Would Bill's body with John's brain be Bill or John? We gather—and hopefully we all are—that personal identity goes with the brain; Bill's body with John's brain would be John rather than Bill. If so, why we take the brain more important for personal identity than (the rest of) the body? The answer is, obviously, as follows: it is because we take the brain as the 'seat' of consciousness (of the person's mental life) and bearer of psychological continuity. As far as we know, a person's psychological continuity is dependent on his-or-her brain rather then on the rest of his-or-her body.

Now, imagine that scientists make a discovery: it is not the whole brain but a small part of it—let us call it 'mind-bearer'—that is responsible for consciousness and psychological continuity. And imagine a possible case of a mind-bearer's swapping between two brains (which remain in their bodies). Would personal identity go with the mind-bearer, or with the body (that retains the largest part of its brain)? We gather—and again, hopefully we all are—that personal identity goes with the mind-bearer, for the same reason that makes us think that in the case of brain-trasplantation, personal identity goes with the brain.

This suggests that what matters for personal diachronic identity is not really the temporal continuity of a body or some its part (supposedly, brain) but psychological continuity. The body's continuity, or the brain's continuity, matters for personal identity only in as much as it bears psychological continuity with it. The materialist seems to be pushed in the direction of the Humean view of personal identity. However, he or she needs not go the full way to the Humean view, with its denial that there are mental subjects as 'havers' of mental states rather then their 'bundles'. Instead, a materialist can think of combining the view that mental subjects are (in fact) brains or (possibly) some other physical mind-bearers with the view that diachronic personal identity consists not in the physical temporal continuity of the brain (or other physical mind-bearer) but in psychological temporal continuity, no matter how it is physically sustained (whether by a physically continuous brain, or by a series of physically discontinuous brains or other

physical mind-bearers). Such a combination is what we call *the quasi-Humean theory of personal identity*.

On this theory, at any moment a mental subject is a human body (brain), but diachronically, a person is not an enduring entity (physical or not) but a series of momentary bodies that are connected not by physical continuity but by relevant psychological continuity. In other words, diachronic personal identity (sameness)—the sameness of a mental subject M that exists and is identical with a body (brain) B at a moment *t* with a mental subject M_0 that existed and was identical with a body (brain) B_0 at a moment t_0—is a matter of temporal *psychological* continuity: there was a continuous process of changes that leads from the set of mental states of B_0 at t_0 to the set of mental states of *B* at *t*—no matter whether B is the same body as B_0 (in the sense of their temporal *physical* continuity) or not. Although at any moment, a mental subject is identical with some body (brain), its criterion of diachronic identity *as a mental subject* is different from its body's criterion of diachronic identity *as a physical body*; consequently, the same subject can, in principle, be identical with different physical bodies (brains) at different times.

To see how this is possible (in principle), consider Derek Parfit's thought experiment (Parfit 1986: 200). Imagine that science has advanced so far that it is now technically possible to create an exact copy of a living human body out of chemicals. On the view we are discussing, you can now travel to some distant planet by undergoing the following process: (1) a device scans detailed information about your body, (2) this information is transmitted to the planet, and (3) simultaneously, your body on Earth is pulverized and exactly the same body is created on the planet. The body created on the planet is not (numerically) physically continuous with your former Earthly body; they are different physical bodies; the new body does not contain even a single atom from your former body. However, it is still you; you are now (identical with) your new body, just as some time ago you were (identical with) your body on Earth.

There is a price to pay for the acceptance of this conception of personal identity: it makes diachronic identity non-transitive relation, which may seem implausible. The transitivity of a relation means that if A is identical with B and B is identical with C, then necessarily A is identical with C. Nevertheless, the quasi-Humean theory of personal identity entails the possibility of a diachronic violation of this relation: a subject S is identical with a body B at time t, and S is identical with a subject S_0 at a time t_0, and S_0 is identical with a body B_0 at t_0; however, B is not identical with B_0. This contradicts the conventional notion of identity as a transitive relation, but perhaps it is expedient to revise the notion. Probably, some materialists would not consider this a too high price. Still, the quasi-Humean theory of per-

sonal identity has also other consequences that are much more implausible and likely to be found unacceptable.

Such consequences are vividly brought out by the thought experiments with which Parfit begins his discussion of the problem of personal identity. The first one is the experiment described two paragraphs above. In such an imagined situation of 'teletransportation', many would find it implausible, or at least very doubtful, that this would be travel rather than murder. If you consider the proposition to travel to a distant city (or planet) not by means of ordinary transport (or spaceship) but by means of your body's being annihilated and qualitatively exactly the same body being created at the destination place (on materialistic presupposition, these bodies will be psychologically continuous, because there is nothing to mental states besides the physical states of the brain), it is likely that you will refuse, because you think, or are afraid, that that body at the destination place will be not you but another person, even if qualitatively the same as you.

Suppose now that you are not travelling at all. You live your ordinary measured life in your city, and do not even suspect that at some moment *t*, scientists created your exact physical (and, hence, on materialistic presuppositions, psychological) copy at some distant place. That person has all your memories, and believes that he-or-she is you that travelled (by means of teletransportation) to that place. Now you get informed of this situation. What would you think of it? We guess that you would think that that person is surely not you, that he-or-she is another person, although at the moment *t*, he-or-she was qualitatively exactly like you in all physical and psychological respects.

A supporter of the *quasi-Humean theory of personal identity* can bite the bullet. Given that he-or-she has already admitted that the relation of personal identity is not diachronically transitive anyway, he-or-she can admit the possibility of branching: a person A at t_0 can be diachronically identical with each of the two persons B_1 and B_2 at t_1, although B_1 and B_2 are not (synchronically) identical (are two different persons). This would give just another violation of diachronic transitivity. However, think of the moment when one of the persons that resulted from the branching, B_1 or B_2, dies (it is very likely that they will not die simultaneously). What should we say of A now? Is he-or-she dead or alive? Or is A both dead and alive simultaneously?

Think now of another modification (similar to the one discussed by Parfit) of this thought experiment. Suppose that you crave to enjoy your summer holidays in the Mediterranean, but you cannot because you have a lot of pressing job to do. Now you are said that you can combine both in an unusual way. On the one hand, you 'travel' teletransportationally to the Mediterranean and enjoy your summer holidays. However (unlike in the

first version of Parfit's thought experiment), your body in your city is not annihilated at that moment but continues to do your job—it will be annihilated only after all the required job is done, just before you return from your Mediterranean holidays. We reckon that if you think of this description, you will probably think that it is a *misdescription*: the right description would be not that you enjoy your holidays in the Mediterranean while your old body toils on your job; it would be that you toil on jour job while your newly created twin enjoys his-or-her holidays in the Mediterranean, and then you are murdered. This seems to be very important; it seems that in such a situation, for a person, the answer to the question 'Which of the two persons after the branching is me?' makes all the difference in the world!

However, Parfit argues that *if it is not the case that you are not a nonphysical entity distinct from your body and brain (a Cartesian Ego*, or *soul),* then your are mistaken to think that these two descriptions are really different, except in wording, so that one of them is true while the other is false. The question, which one of the two description is true (equivalent to the question 'Which of the two persons after the branching is me?'), is *an empty question*; there is no true answer; we can choose to describe the situation in either way. The answer can only be arbitrary and makes no difference, because there is no real fact behind it. All there is to the situation is that a new body is created at some moment t_0 and the old body is annihilated at a later moment t_1, and that this results in the branching of psychological continuity at t_0 and one of the branches being cut at t_1. And of this, all that matters, according to Parfit, is psychological continuity; *personal identity is indeterminate and does not matter*. The same should be true also for all other cases, including those in which personal identity is not problematic. *If we do not believe that we are Cartesian Egos*, we should conclude that all that matters in the vicinity of personal identity is psychological continuity (personal identity matters only insofar as it involves psychological continuity, or marginally a bit more or less).

One can think that this Parfit's conclusion is due to the assumption of the quasi-Humean theory of personal identity but can be avoided if we accept the body theory or the brain theory of personal identity. Parfit's other thought experiments, with body-and-brain-continuity-spectrum and with brain-branching show that this is not so.

Consider the body-and-brain-continuity-spectrum thought experiment (Parfit 1986: 236-237). Think again of Bill and John. Suppose John has died as a result of an accident, and suppose that science is so developed that it is possible to recreate John's body out of atoms, alive and healthy, just as it was before the accident. Imagine now two possible events that seem quite unconnected: (1.1) a tiny change has happened with Bill's body and brain, of the scale that happens continuously with each of us, and (2.1) Bill's body is annihilated and John is 'resurrected' by scientists (his body is recreated,

alive and unimpaired). Imagine now another two possible events: (1.2) there happened a tiny bit more considerable and quicker change with Bill's body and brain, in a direction that make Bill just a tiny bit more like John than he was before, both physically and psychologically; (2.2) John was 'resurrected' by scientists with a tiny difference in his body and brain, in a direction that make John just a tiny bit more like Bill than he was before, both physically and psychologically. Now we can think of further changes in these directions, so that changes with Bill became more and more abrupt and he becomes more and more like John; at the far end, this will be the same as (2.1), so we should say that the resulting body and person is John's, not Bill's. Analogously with John: think of the situations when Bill's body is annihilated, and 'John's' body is recreated with more and more changes that make him more and more like Bill, even with atoms taken from Bill's body; at the far end, this is the same as to annihilate and then recreate Bill's body, alive and healthy, so we should say that the resulting body and person is Bill's, not John's. Now think of the situation just in the middle of the spectrum, when the resulting body and brain (and, hence, mental constitution, memories, temperament etc.) are equidistant from initial Bill and initial John. In this situation, is Bill dead and John alive, or the other way round? Or consider to marginally different cases A and B very close together in the middle of the spectrum; they are almost indistinguishable, but there is a slight difference: the case A is just a tiny bit more like initial Bill than initial John, while the case B is just a tiny bit more like initial John than initial Bill. Would it be right to say that in the case A, Bill is alive and John is dead, whereas in the case B, John is alive and Bill is dead? This cannot be so consistently with materialistic assumptions, because the cases A and B are only marginally different, and this marginal difference cannot have such radical consequences as Bill's and John's lives and deaths; A and B are so close together (and much more distant to initial John and initial Bill) that if some person is alive in these cases, it is surely the same person. From all this, we should draw (consistently with materialistic assumptions) the conclusion that personal identity is not a matter of 'either-or' ('all-or-nothing') but a matter of degrees, so that in some cases that can conceivably happen with me (imagine yourself in Bill's place), it is indeterminate whether the surviving person is me or not me. In such cases, the question 'Do I survive or die?' (which naturally seems to make all the difference in the world) is, according to Parfit, an empty question. Alternatively, on the Cartesian assumption that we are immaterial indivisible entities, or souls, all would depend on which soul, Bill's or John's is associated with a body.

We may arrive at the same conclusions by considering the possibilities of brain-branching. There are persons who survive with only half of the brain. Imagine two halves of a person's brain transplanted into two bodies that

survive as two different persons. Or imagine some (really non-existing but conceivable) process in which each cell of a human brain divides into two structurally exactly the same cells, absorbs the lacking stuff from the outside, and then all these cells get arranged into two brains with all the connections exactly the same as in the initial brain. On materialistic assumptions, this should ensure enough physical continuity, and almost perfect psychological continuity. Again, the question 'Which one of the two resulting persons is me?' is, according to Parfit, an empty question; personal identity is indeterminate and does not matter; all that matters is psychological continuity.

Should we really accept such implausible consequences? Parfit argues that we should, because the only coherent alternative is the belief that we are non-physical entities, distinct from our bodies and brains—Cartesian Egos, or souls. What is supposed to be wrong with that belief? Parfit advances several objections.

The Cartesian View on Trial

(1) The *no-reason-to-believe-in-the-existence objection*. It has two parts (premises):

(1.1) We have no evidence that such non-physical entities (souls, or Cartesian Egos) do exist, or that a person is such a separately existing entity.

(1.2) 'if we have no reasons to believe that such entities exist, we should reject this belief. … My claim is merely like the claim that, since we have no reason to believe that water-nymphs or unicorns exist, we should reject these beliefs.' (Parfit 1986: 224)

We think that this objection is straightforwardly unfair. Unlike in the cases of water-nymphs or unicorns, we do have reason to believe that we, as mental subjects, are non-physical entities distinct from our bodies and brains and, therefore, that such entities exist. The reason is just that our bodies and brains are not mental subjects; they are merely very complex physical systems—complexly ordered huge multitudes of microphysical entities (such as atoms, or electrons) that interact according to the laws of physics; none of these constituents have any subjective experiences (mental states), and there is nothing to our bodies (brain) besides these constituents and their physical-laws-abiding interactions. There is nothing subjective about this; there are no mental subjects in the picture.

There are well-known and widely discussed arguments against materialism that are best interpreted as devices to highlight this point, for instance Leibniz' 'mill argument' (Leibniz 1965: 150); Kripke's argument (Kripke

1972: 334-342); the zombie argument (Kirk 1974a; Kirk 1974b; Chalmers 1996: 94-99); the knowledge argument (Jackson 1982; Robinson 1982: 4-5; for a more recent defence, see Fumerton 2013; Robinson 2016). If construed in a different way, as deductive arguments in which the conclusion validly follows from independent premises, these arguments can be charged with begging the question: it can be said that they assume that mental states are not (constituted by) some physical processes in our brains or functions those physical processes perform). Regrettably, Parfit does not mention the existence of such anti-materialistic arguments.

It can be noted that in the contemporary philosophy of mind, anti-materialistic arguments, such as the knowledge arguments and the zombie argument, are usually advanced and discussed as arguments for the existence of some non-material (non-physical) mental states (*qualia*) rather than of non-material mental subjects. We consider this as a manifestation of the influence of the Humean view. If this view is discarded, it should be obvious that the existence of non-material mental states implies the existence of non-material mental subjects.

(2) The *no-evidence-for-the-continuity objection*

> Suppose that I was aware that I was such an entity... I could not know that this entity continued to exist. As both Locke and Kant argued, there might be a series of such entities that were psychologically continuous. (Parfit 1986: 223)

To estimate the strength of this objection, we propose to compare two hypotheses. The first one is the well-known Kant's hypothesis (Kant 1964: 342) appealed to by Parfit:

(KH) For every human person, there is a series of psychologically continuous mental subjects rather than one continuous mental subject

The second hypothesis was proposed by Bertrand Russell (Russell 1921: 159):

(RH) The world did not exist for any long time. It sprang into being (out of nothing) just five minutes ago, with ourselves and all our memories, and with all that we may take as traces of past events.

Now, just as with (KH), we cannot know that (RH) is not the case. If it were the case, everything would be for us just as it is in fact. Russell's hypothesis (as well as lots of other skeptical hypotheses) cannot be refuted by any con-

ceivable evidences. No possible evidence can distinguish between the possibility envisaged by Russell's hypothesis and what we take to be really the case. However, no one takes this seriously as a good reason to abandon the belief that the world exists for a long time. We are of the opinion that our attitude toward Kant's hypothesis should be the same.

In may be worth noting that the two hypotheses are not just superficially similar but deeply connected. To be more precise, there is a deep connection between their commonsense opposites: the commonsense view (NKH) that we are enduring mental subjects (for example, my belief that it was me—not someone else—who lived a day ago, a year ago, and thirty years ago and had those experiences that I remember as mine) and the commonsense view (NRH) that the world exists for a long time rather than has emerged out of nothing just a moment ago. (NKH) grounds (NRH); without (NKH), there is no reason to hold (NRH). To see this, note that Kant's hypothesis involves two distinct assumptions:

(1) I (you) as a mental subject emerged just a moment ago with all memories ready.
(2) There was a series of fleeting mental subjects that are sort of mental ancestors of mine—from which I inherited my memories. And there was the world that was the source of these subject's experiences and what seems to be my memories.

However, if (1) is the case, then all my seeming memories are fictitious quasi-memories. (It is so, because what seems to be my memories are as of *my* experiences and *my* interactions with the world; however, according to (1), I could not have those experiences and interactions, because I did not exist.) If so, there is no reason at all to believe in (2). (1) is far more congenial with the hypothesis that there was no my mental ancestors and no world—the world has emerged just a moment ago, together with me and my quasi-memories, just as Russell's hypothesis envisages.

(3) *The psychological irrelevance objection*

(3.1) We do not have evidence to believe that psychological continuity depends chiefly on the continuity of some other entity rather then of the brain, and we have much reason to believe both
(3.2) that the carrier of psychological continuity is the brain, and

a. that psychological connectedness could hold to any reduced degree (Parfit 1986: 228).

Although all the premises of the objection seem true in a sense, it does not follow from them that the Cartesian view of ourselves is false, or that personal identity, like psychological connectedness, 'could hold to any reduced degree', or that it does not matter. It should be admitted that the objection has *some* force against the Cartesian view of ourselves, but the Cartesian view can withstand that force.

The force of the objection is due to the fact that for our idea of ourselves, psychological connectedness does matter. However, there are two reasons why that force is not enough to defeat the Cartesian view:

(1) Psychological connectedness is not the only thing that matter for our idea of ourselves; there is another constituent of this idea that matters very much and more fundamentally—the (Cartesian) notion of ourselves as unitary mental subjects.

(2) All the available evidence for the role of the brain in the maintenance of psychological continuity falls short of establishing that the brain is *the* carrier of psychological continuity, that physical continuity of the brain is both necessary *and sufficient* for psychological continuity, that no other (non-physical) entity plays its role in the maintenance of mental states and their continuity.

To make a case for (1), let us suppose that despite (2) the brain is the carrier of psychological continuity, and of the whole psychological profile of a person. If that were the case, how much would the Cartesian view suffer?

We think that in such a case, the arguments for a version of the Cartesian view still hold: the body (brain) as a physical system is not a mental subject; the mental subject is a non-physical entity. However, if the supposition that the brain determines the whole psychological profile of a person is true, this means that all these (quasi-)Cartesian Egos are intrinsically qualitatively the same; all qualitative differences between the mental states of numerically different mental subjects are entirely due to physical differences between their brains.

On this view, if Hitler's and Mahatma Gandhi's bodies (brains) swapped their associated souls, nothing would change for the world—the person with Hitler's body and brain and Gandhi's soul would behave exactly like Hitler did, and analogously for Gandhi. (Of course, no Cartesian really believes such a thing.) In this case, even if we knew about the fact of Hitler-Gandhi soul-swapping, we would think of the person with Hitler's body and brain and Gandhi's soul as Hitler rather that Gandhi. This seems to support the view that psychological continuity matters more for our notion of a person than a mental subject's (considered as an entity with no individual psychological profile) continued existence. However, we propose that this im-

pression is misleading. It reveals only that what matters to us primarily about most of other persons is the role that they play in our lives; we are not much interested in those persons (as mental subjects) for their own sake. In relations with most of other persons, we naturally take an extrinsic attitude, from outside, and this attitude is concerned only with what can we expect from a person and, hence, with his-or-her psychological profile.

However, with respect to ourselves and those persons who are dear to us for their own sake (rather than for what they do for us), our way of thinking is likely to be different, in the way that shows that in our notion of ourselves, the Cartesian constituent matters very much. Suppose you can imagine the possibility of an accident in which you lose all your memories and there are injuries to your brain that make your temperament different beyond recognition, but still think of that person as yourself. If you are a Cartesian, you have a theoretical ground for this. If you are not, supposedly you still can imagine the situation; the description does not seem senseless, or incoherent. In an earlier philosophical dialogue with Godfrey Vesey, Parfit remarked that even if we do not realise it, we are all inclined to hold '[t]he belief that however much we change, there's a profound sense in which the changed us is going to be just as much us', '[t]hat even if some magic wand turned me into a completely different sort of person—a prince with totally different character, mental powers—it would be just as much me' (Parfit and Vesey 1974).

So, as Parfit admits, we all (whether Cartesians or not) are naturally inclined to believe that there is a 'deep fact' of our endurance as mental subjects, which can conceivably part from psychological continuity. At some level, implicitly, we all (even Parfit, as he admits himself) believe it, even if we explicitly reject that belief when philosophizing (as Parfit does). That implicit belief permeates the whole bulk of our everyday reasoning and, what is the most important, our valuations—what fundamentally matters for us, or what we take as what fundamentally matters (in the sense of ultimate value, as distinct from means for something other).

In that valuation, our endurance as mental subjects is what matters for its own sake, not for the sake of psychological continuity. Moreover, in that valuation, psychological continuity matters not for its own sake, but for its *being ours*. To see this, let us modify the example with Gandhi and Hitler. Suppose it is not Gandhi's soul that is going to undergo the change but your own soul, or soul of a person who is dear to you for his-or-her own sake (not just for relationship with you). Would you be indifferent to the prospect of such a soul-swapping? You would probably not like the idea that tomorrow you will find yourself in Hitler's body, thinking of yourself as Hitler, and having all Hitler's memories, beliefs, valuations and temper, even if your former psychological stream continues with Hitler's former

soul that inhabits your former body-and-brain. Or imagine that there is a person who is going to undergo terrible tortures tomorrow, and there is a magician who is going to move your dear child's soul into that person's body and make it perfectly psychologically continuous with that person (as he-or-she is now), while moving that person's soul into your child's body and making it perfectly psychologically continuous with your child (as he-or-she is now). Would you take these magician's manipulations as things that do not matter?

Really, we think that most of us would not agree to swap our (or our dear one's) souls with another person's body, if this involves the loss of *our* (or our dear one's) psychological continuity, even if that person is not such an abominable creature as Hitler and not an unhappy man to be tortured but a good-hearted and wealthy and happy human being. Psychological continuity does matter much for us, but it is *our* psychological continuity that matters, not just psychological continuity. In all those imagined cases of soul-swapping mere psychological continuity is retained: both psychological streams continue, but they continue with other selves, and that is exactly what makes a difference. What matters for me, so far as psychological continuity is concerned, is not that there was a psychological stream qualitatively continuous with my present psychological stream, but that *my* psychological stream continues with *me* (the same about my dear ones).

Perhaps, Parfit would agree that this is *what we (are naturally inclined to) take as what matters* but insist that this is *not what really matters*, and that what really matters is mere psychological continuity, because there is no 'deeper fact' to account for the difference between *mere* psychological continuity and *my* psychological continuity. A Cartesian reply is twofold: (1) there is such a deeper fact: it is the fact of the mental subject's enduring existence (which is possible if a mental subject is an irreducible entity, like a Cartesian ego, or soul) and (2) if that fact is denied, the natural ground for the valuation of a person's psychological continuity (which is the valuation of a certain mental subject's—my or your—psychological continuity) is lost, and Parfit's claim that mere psychological continuity is what really matters is quite arbitrary, or only marginally not so. By marginal mattering we mean the following. Suppose, you know that somewhere in the Universe, or in a parallel world, there is a person who is psychologically very much like you. How much would you naturally worry about that person's survival and well-being? You would probably worry only marginally if at all—surely, far less than about your own and your dear ones' survival and well-being.

So far, the discussion of Parfit's *psychological irrelevance objection* proceeded on the supposition that is most unfavourable for the Cartesian view: that a person's psychological profile, and its continuity, is entirely a matter of brain structures; that the Cartesian ego (or soul), even if it exists, makes no

person-specific contribution to the psychological profile. The preceding argument demonstrated that even on that supposition, what matters for a person is *his-or-her* (as a mental subject's) continued existence and *his-or-her* psychological continuity, rather than *someone's* psychological continuity with *him-or-her*. On the other hand, it should be admitted that the supposition at issue (that our psychological profile is entirely determined by our brains, and not at all by our souls) can detract very much from our natural valuation of our (as a mental subject's) continued existence (especially in certain contexts, such as considering the hypothesis that after my bodily death, my soul will continue its existence in another body).

The continued existence of our souls is likely to seem much more valuable if we believe that our psychological profile depends (if not entirely, then at least to a considerable degree) on our souls, especially if we believe that our souls somehow retain imprints (if not explicit memories) of our present life experiences and developments. Such a belief is natural for a person who believes in the existence of the soul, and he-or-she surely should not abdicate it without a very strong reason. Is there such a reason? Parfit claimed that there is: 'we have much reason to believe that the carrier of psychological continuity is the brain' (Parfit 1986: 238).

However, what kind of reason that is? What exactly it establishes about the brain's role in the maintenance of psychological continuity? There is much evidence of a negative kind: that the brain maintains psychological continuity in the sense that when certain brain areas are damaged, psychological continuity (as manifested in behaviour) suffers. Such negative evidence is enough to establish that, insofar as our earthly experience allows us to judge, the brain's physical continuity is a *necessary* condition for psychological continuity. However, that does nothing to establish that the brain's physical continuity is the *sufficient* condition for psychological continuity, that nothing else (such as a non-material soul) plays its role in the maintenance of psychological continuity. For such a stronger claim, which could refute the belief that a soul matters for psychological profile and continuity, one would need much more radical, positive evidence, such as creating a mature person's exact physical copy out of atoms. (If that creature turns out to be an exact *psychological* copy of the initial person, then Parfit's case is made!) At present, science cannot provide such evidence, and we do not expect that it will ever do.

(4) *The objection from the cases of brain bisection*

The most severe cases of epilepsy, a neurological decease characterised by seizures, were often treated by surgical severance of the system of fibres called *corpus callosum*, which connects the right and the left hemispheres of

the brain core. The severance efficiently decreased the number of seizures and had no grave negative effects.

There is some division of data processing and body-control job between the two hemispheres: they are sort of responsible for data input and behavioural output each of one half of the body. However, with normal people (whose *corpus callosum* is unimpaired) the work of two hemispheres is perfectly coordinated. With brain-bisected patients, in ordinary situations, it seemed that the coordination is retained (there are some other, more indirect links between the hemispheres besides the *corpus callosum* that can account for this coordination). However, there were experiments in which scientists neatly isolated input data and tasks for the two hemispheres: each hemisphere has no access to the information and tasks provided for another hemisphere. In such situations, it turned out that the behaviour of brain-bisected patients—or rather, or the left and the right halves of their bodies—lost its coordination and strongly suggested that there are two distinct conscious persons associated with the left and the right hemispheres. It seemed that each of these 'subpersons' has its own experiences and emotions, both have memories of the initial person and identify themselves with it. (If the initial person is Smith, then each 'subperson' is capable to communicate that his-or-her name is Smith; each is capable to recognize Smith's relatives, etc.)

There is a strong difference in the capabilities of these two 'subpersons': the left hemisphere's 'subperson' is far superior to the right one's. In particular, the left hemisphere is normally responsible for speech and abstract thinking; so the left hemisphere's 'subperson' is as good a speaker as the initial (normal) person; whereas the right hemisphere's 'subperson' is at best capable to identify single words or simple phrases with objects or pictures and perhaps write such words or simple phrases. It seems that the higher functions of language—descriptive (formulation of statements to describe situations) and argumentative functions are beyond the ken of the right hemisphere's 'subperson' (on the relevant functions of language, see Popper 1963; Popper and Eccles 1977: 57-59). Generally, as far as we gather from various sources, the performance of the right hemisphere's 'subperson' seems to be nearer (although in some respects superior) to the performance of the cleverest non-human animals (such as chimpanzees) than of normal human persons.

In the experimental situations at issue, each 'subperson' is ignorant of what the other one experiences and does. After the experiment, if a patient is asked what he-or-she did in the experiment, he-or-she will report the experiences and actions that pertained to the left hemisphere; he-or-she cannot recollect experiencing and doing what pertained to the right hemisphere.

What conclusions should we make out of this? Different authors, including the Noble Prize Winners in neurophysiology Robert Sperry and John Eccles, made radically different, almost opposite conclusions, and their arrival point seems to perfectly reflect their starting point (Gazzaniga 1971; Nagel 1971; Popper and Eccles 1977: 311-333; Sperry 1977). Most authors, including Parfit, who belong to the materialistic mainstream, take the results of the experiments as a refutation of the absolute unity of consciousness, or of the mental subject. Other authors, such as John Eccles, take the same results as a strong corroboration of the absolute unity of the mental subject: the mental subject associated with the left hemisphere remains perfectly continuous through all the described peripetia.

As for the right hemisphere, there are two possibilities consistent with the Cartesian view (or something near enough):

(1) There is no real mental subject associated with the right hemisphere. Its performance is a very good automatic 'imitation' of a conscious human person's behaviour. Some suggestions by Wilder Penfield (Penfield 1975: 37-59) and Karl Popper (Popper 1974: 152-153) may be relevant to this point. Our brains and bodies are capable of performing very sophisticated activities while we are not conscious or barely conscious of performing them. We can drive a car, or play a piano, or go to some place by a tortuous route without being aware, for the largest part, of what we are doing in these respects; we may think of other things and only episodically (and that when something goes out of routine) pay attention to driving, piano-playing, or going. Learning these performances took much conscious attention and effort; however, when the study was successfully completed, those (sometimes very sophisticated) activities became 'programmed' somewhere in the brain and do not require consciousness any more, except when something goes wrong. The marvelous performances of the right hemisphere may be the same kind of highly sophisticated *unconscious* (although seeming very conscious-like) automatisms.

This view raises an interesting question about higher animals: if their brain's performance is not superior to the human right hemispheres' performance, and we think that the right hemisphere has no distinct mental subject associated with it, should we think that higher animals, such as dogs or chimpanzees, are not mental subjects but mere automatons (as Descartes believed), that is, that they have no subjective mental states (experiences); there is nothing it is like for them to have a pain, or to see a red thing? We suggest that there is reason not to reach this conclusion. The reason is that with higher animals' brains (unlike the right hemisphere of the human brain) there is no higher consciousness-associated system (such as the left

hemisphere of the human brain) to 'program' them; insofar as it is not genetically inherited, their performance is their own achievement.

(2) There is a real (although inferior) mental subject (soul) associated with the right hemisphere. If so, this in no way infringes upon the enduring existence and self-identity over time of the superior mental subject (soul), which is associated with the left hemisphere. It may be that in normal human people (with *corpus callosum* unimpaired) the brain functions in a way that make mental states of the two mental subjects systematically cohere, whereas with brain-split patients, the coherence may break in some situations.

Parfit advances a further argument based not on known facts but on what he thinks conceivable about brain-bisection-like cases. He claims that the following possibility is conceivable: at some moment his psychological stream splits into two streams (associated with the left and the right hemispheres of the brain core), so that each is psychologically continuous with the initial stream, and each has its own experiences, performs its own data-processing task, and is ignorant of what the other stream experiences and does; after a spell, the streams reunite, so that the re-united stream is psychologically continuous with both branch-streams and (through them) with the initial stream (Parfit 1986: 246-247). Conceivably, in that re-united state, Parfit has his pre-split memories plus the memories inherited from both branch-streams: he remembers himself having the left hemisphere's experiences and performing its data-processing while being ignorant of the right hemisphere's experiences and data-processing and, *simultaneously*, having the right hemisphere's experiences and performing its data-processing while being ignorant of the left hemisphere's experiences and data-processing.

What is the relevance of this thought experiment to the evaluation of the Cartesian view? We think that the gist is as follows. If the described situation is conceivable, this would mean that we can make sense of the idea that mental subjects can be indeterminate. This would detract very much from the attractiveness of the Cartesian view, because it owes very much to the apparent nonsensicality of the subject-indeterminacy view. However, *pace* Parfit, it is far from clear that the described situation is conceivable. We think it is not: the description of what Parfit remembers after the re-union seems to me incoherent. I, for one, cannot conceive of myself as (being aware of) having an experience A and not having an experience B (or being ignorant of having it) and, *at the same time* (being aware of) having an experience B and not having an experience A (or being ignorant of having it). We expect that this description is clearly incoherent. And because such a mental state is inconceivable (its description is incoherent), remembering such a state is also inconceivable.

Conclusions

Parfit succeeds to make a strong case that there is no plausible and coherent way to reconcile materialism (or something near enough) with the view of personal identity as necessarily determinate, to which we are all very much inclined and which seems necessary to make sense of the first-personal perspective. This leaves us with the alternative: either we take the view Parfit calls 'Reductionism', which is radically at odds with our natural take of ourselves and glaringly fails to accommodate the first-person perspective, or we take the substance dualist view (or something near enough) that we (as mental subjects) are ontologically fundamental non-material mental entities (Cartesian egos, or souls). Substance dualists can be grateful to Parfit for this illumination. However, Parfit's evaluation of the two alternatives is biased. Although he honestly admits many implausible consequences of Reductionism, Parfit underestimates its values-undermining effects: if what we naturally take as what matters (*a person's* continued existence and *his-or-her* psychological continuity) is discarded, and no reason is provided to believe that *mere* psychological continuity matters *on its own* (no matter whether it is me or not me that is psychologically continuous with myself), then the conclusion seems to be that nothing matters rather than that *mere* psychological continuity matters. Parfit actually tried to avoid this conclusion by claiming that there is no weighty argument for either side; however, this ignores the obvious consideration as to on which side the burden of proof lies. To remind Parfit's other claim: 'since we have no reason to believe that water-nymphs or unicorns exist, we should reject these beliefs'. One need only replace the question 'What exists?' with the question 'What matters?' On the other hand, Parfit's arguments against the Cartesian view (or something near enough) are far weaker than he takes them to be. On balance, soul-believing looks reasonable.

Bibliography

Chalmers D (1996) *The Conscious Mind*. New York, NY: Oxford University Press.

Fumerton R (2013) *Knowledge, Thought, and the Case for Dualism*. Cambridge: Cambridge University Press.

Gazzaniga M (1971) The split brain in man. In Thompson R (ed) *Physiological psychology*. San Francisco, CA: Freeman, pp. 118-123.

Hume D (1951) *A Treatise of Human Nature*. Oxford: Oxford University Press.

Jackson F (1982) Epiphenomenal Qualia. *Philosophical Quarterly* 32(*): 127-136.

Kant I (1964) *Critique of Pure Reason*. London: Macmillan.

Kirk R (1974a) Sentience and Behaviour. *Mind* 83(329): 43-60.
Kirk R (1974b) Zombies v. Materialists. *Proceedings of Aristotelian Society* 48(*): 135-152.
Kripke S (1972) Naming and Necessity. In Davidson D and Harman G (eds) *Semantics of Natural Language*. Dordrecht: Reidel, pp. 253-355.
Law S (2002) Where am I. In Law S (ed) *Philosophy Files*. Orion Children's Books, pp. 55-77.
Leibniz W (1965) *Monadology and Other Philosophical Essays*. New York, NY: Bobbs-Merrill.
Lichtenberg G (1971) *Schriften und Briefe*, volume 2. München: Carl Hanser Verlag.
Nagel T (1971) Brain Bisection and the Unity of Consciousness. *Synthese* 22(*): 396-413.
Parfit D (1986) *Reasons and Persons*. Oxford: Oxford University Press.
Parfit D and Vesey G (1974) Brain Transplants and Personal Identity. In Vesey G (ed) *Philosophy in the Open*. Open University Press, pp. 54-65.
Penfield W (1975) *The Mystery of the Mind*. Princeton, NJ: Princeton University Press.
Popper K. (1963) Language and the Body-Mind Problem. In Popper K. *Conjectures and Refutations*. Routledge & Kegan Paul, pp. 293-299.
Popper K (1974) Intellectual Autobiography. In Schilpp P (ed) *The Philosophy of Karl Popper*, book I. La Salle, IL: Open Court, pp. 3-181.
Popper K and Eccles J (1977) *The Self and Its Brain*. New York, NY: Springer.
Robinson H (1982) *Matter and Sense*. Cambridge: Cambridge University Press.
Robinson H (2016) *From the Knowledge Argument to Mental Substance*. Cambridge: Cambridge University Press.
Russell B (1921) *Analysis of Mind*. London: Allen & Unwin.
Russell B (1945) *A History of Western Philosophy*. New York, NY: Simon and Schuster.
Sperry R (1977) Forebrain commissurotomy and conscious awareness. *The Journal of Medical Philosophy* 2(*): 101-126.