

The Argument for Subject-Body Dualism from Transtemporal Identity

KIRK LUDWIG

Indiana University

1. Introduction

In recent papers, Martine Nida-Rümelin (NR) has argued for a position she calls subject-body dualism on the basis of three distinct arguments.¹ One is based on reflection on the emergence of phenomenal consciousness (Nida-Rümelin 2010b). One is based on reflection on our treatment of conscious subjects as active (Nida-Rümelin 2006). One is based on reflection on intelligible possibilities of transtemporal identity in fission cases drawn from the literature on person-identity over time (Nida-Rümelin 2010a). In this paper, I am concerned only with last of these arguments. Arguments from reflections on personal identity to a form of subject-body dualism have been around for a while. For example, (Swinburne 1984, 1986) has focused on the implications of duplication cases as well as the possibility of survival of the destruction of the body. More recently (Lowe 2010a, sec. 1.4) has argued for a form of subject-body dualism on the grounds of a difference between the identity conditions for persons and bodies. What I take to be distinctive about NR's argument is the focus on there being a factual difference between the claims that the original subject is one, or the other, of the two resultant subjects in fission cases. It is the role of this claim in her argument that will be my central focus. I argue that on each of the three most plausible interpretations of this assumption the argument fails.

¹ The general idea that subjects of experience are emergent individuals not identical to their bodies has been advanced in recent times by a number of authors. See for example (Hasker 1999; O'Connor and Jacobs 2003; Nida-Rümelin 2007; Lowe 2010b, 2010a; Zimmerman and Mackie 2010).

By a subject, NR has in mind a subject of conscious experience that persists through time while its properties change. Thus, NR rejects four-dimensionalism about subjects of experience. She distinguishes this from traditional substance dualism, which holds that there are distinct and independent substances that are the bearers of mental and physical properties, respectively. The distinction rests on the denial of the claim that subject-body dualism is committed to holding that the subject of experience can exist without a body. On Descartes's account of substance dualism, not only were mental and physical substances distinct, they were necessarily so. Given the form of the argument for subject-body dualism, I think it is likewise committed to the thesis that mental and physical substances are necessarily distinct, if not independent. If subject-body dualism can be established on the grounds NR advances, then ontic materialism is not only false but also necessarily false.²

Subject-body dualism holds that the subject of experience is not (i) identical with, or (ii) constituted by any material body, (iii) that it is not composed even in part by a body, (iv) that it is not, on the contrary, composed of some immaterial stuff, (v) nor an immaterial part of a person, and (vi) that it has a location only in the derivative sense of being where its body is.³ The thesis, being grounded in an argument that involves reflection on being a subject of conscious experience, extends beyond humans to all creatures capable of (phenomenal) consciousness, e.g., to all creatures capable for feeling warmth or pain.

It is, I think, antecedently implausible that it is impossible that subjects of experiences be material objects located in space-time. Either ontic materialism is true or it is false. If it is true, then it is not impossible that subjects of experience be material objects located in space-time. If it is false, then some form of property dualism is true. But then, as Spinoza observed in his criticism of Descartes's version of substance dualism, as the attributes of thinking and of extension are absolutely independent of one another, there is nothing to prevent one

² I will use 'ontic materialism' to denote the view that every object, event or process is physical. I will count a thing as physical if it is constituted by a physical thing or by physical things (and nothing else). I will use 'reductive materialism' to denote the view that mental properties conceptually supervene on the physical. Ontic materialism is incompatible with subject-body dualism, but not with property dualism, while reductive materialism is incompatible with property dualism, but compatible with, for example, functionalism.

³ Points (ii) and (vi) distinguish NR's position from that of (Baker 2000) and (Lowe 1996, 2006); points (iii) and (iv) from (Swinburne 1986).

substance having both (Spinoza 1994, p. 90; IP10(Schol.)). Still: each argument for a form of ontic dualism must be addressed on its own terms.

NR's master argument has three premises:

1. There is a factual difference between the claim that someone is one or the other of two continuers in fission cases or we are subject to a pervasive illusion in our thoughts about personal identity over time (the illusion theory).
2. There could be a factual difference between the claim that someone is one or the other of two continuers in fission cases only if subject-body dualism were true.
3. The illusion theory is untenable.
4. Subject-body dualism is true.

I will grant premise 3 for the sake of argument. We can then reduce the argument to the following form.

1. There is a factual difference between the claim that someone is one or the other of two continuers in fission cases.
2. There could be a factual difference between the claim that someone is one or the other of two continuers in fission cases only if subject-body dualism were true.
3. Subject-body dualism is true.

I will be concerned with the grounds for the second premise of this argument. I consider three interpretations of the claim that there is a factual difference between the claims that one or the other of two equally good continuers of a person in a fission case is identical with her. The first is that one or the other would in fact be identical with the original. This would secure the second premise, but it begs the question, as a *prima facie* equally intelligible option is that the original does not survive. The second is that the two claims express different propositions. I will argue, however, that there is no good reason to think this is incompatible with a materialist position. The third is that in contrast to the materialist position, it is at least possible that one or the other be the original. I will argue that this must be reduced to the claim that something of the type *subject of experience* could survive as

one or the other, but that this yields at best an epistemic possibility that, as things are, subjects of experience survive as one or the other, and from this subject-body dualism does not follow. I will also suggest, however, that if the premise on this last interpretation is correct, a version, not of substance, but of property dualism does follow. If this is right, then while an ontic materialist need not reject the premise, a reductive materialist must either deny that we are subjects of experience (i.e., must embrace the illusion theory and so eliminativism about subjects of experience) or deny even the possibility of survival of subjects of experience in fission cases.

In section 2, I explain the argument for subject-body dualism in enough detail to clarify the target of evaluation. In section 3, I consider each of the three interpretations of the central claim sketched above and argue that on each the claim is inadequately supported or compatible with ontic materialism. In section 4, I show that the assumptions that go into the final interpretation establish property dualism.

2. The Argument from Transtemporal Identity for Subject-Body Dualism

The argument goes as follows. First, we imagine a case in which an embodied person, Andrea, undergoes an operation that produces two equally good candidates for her successor. We may think of a brain that has developed into two identical hemispheres each capable of functioning independently when separated, which are then each transplanted from the original body, which is destroyed in the process, into identical bodies grown for the purpose. I will call such operations *fissioning*. What is required for the argument is not that fissioning be biologically or technically possible as things stand, or even in the limit of human technological development, but just the conceptual possibility, for the argument aims to be getting at something about our concept of a subject of experience, which can be used to draw conclusions about actual subjects of experience whether or not any of them are ever involved in the sorts of scenarios considered.

The two successor bodies we call the L-body and the R-body (for ‘left’ and ‘right’) and the two successor subjects of experience we call L-Andrea and R-Andrea. For whatever empirical relations there are that obtain between Andrea at t and L-Andrea at t' that might be cited to ground the claim that Andrea is identical with L-Andrea, there are symmetrical empirical relations that obtain between Andrea and R-Andrea. These include both physical and psychological facts—excepting psychological facts which would themselves presuppose an answer to the question which of L-Andrea and R-Andrea were identical with

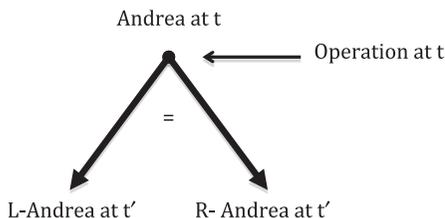


Diagram 1. Symmetrical physical and psychological relations

Andrea—for example, that one of them remembered doing something that Andrea did in the sense of remembering that requires veridicality. This would settle the matter, but not ground it, since it presupposes identity. We are to imagine the case in such a way that if not for the existence of R-Andrea, it would be natural to treat L-Andrea as Andrea, and vice versa. That is, in the absence of the other, each would be not just the best continuer of Andrea but a compelling candidate for being Andrea.⁴ We will call cases that have these features duplication cases. Diagram 1 illustrates the relevant structural features of the scenario.

Let ‘D’ stand in for a description of all the relevant facts of the case independently of specifying the relevant facts about identity (one can throw in as much as one thinks relevant). There are (at least) three possibilities with respect the question with which of L-Andrea and R-Andrea our original Andrea is identical.

P1: D and Andrea is L-Andrea.

P2: D and Andrea is R-Andrea.

P3: D and Andrea is not L-Andrea and Andrea is not R-Andrea.

With NR, I put aside the possibility Andrea survives with two bodies,⁵ that is, I assume that L-Andrea and R-Andrea are distinct persons, and that no one person is identical with two distinct persons.

⁴ I do not intend, however, to take on any commitment as to whether in fact Andrea would be L- or R-Andrea if the other did not exist.

⁵ I will *not* distinguish between Andrea surviving the operation and there being someone subsequent to it who is Andrea.

The first premise of the argument is that there is a factual difference between P1 and P2.⁶ What this comes to is crucial to understanding the argument. It is natural, at a first pass, to take this to mean that what is expressed by ‘Andrea is L-Andrea’ and ‘Andrea is R-Andrea’ are distinct propositions, which would be made true by distinct states of affairs. If the first is true, then Andrea’s experiences after *t*’ are L-Andrea’s and not R-Andrea’s. If L-Andrea feels pain, it is Andrea who feels pain; but if R-Andrea feels pain, Andrea does not. *Mutatis mutandis* if the second of these propositions is true instead of the first. It seems easy to see the difference if one is imagining that it is oneself who will be undergoing the operation. If one wakes up afterwards, it will either be in the one body or the other. When one looks in the mirror, it will be the face of the R-body or the L-body that one sees. When one pinches oneself, one will be pinching the R-body or the L-body, and the pain one feels will be the pain associated with the pinch in the one body or the other. And so on.

Reflection on the perspective of the person facing the operation leads NR to several additional claims.

The first is that future directed self-attributions of properties in the first person (attributions we would express by saying “I will have the property P,” for example, “I will have a headache tomorrow morning”) are conceptually prior to transtemporal self-identifications. That is, we don’t first decide that we are going to be a certain person in the future and then infer that we will have the properties that that person has

⁶ NR puts it this way (Nida-Rümelin, 2010a, pp. 195–6; unless otherwise indicated, henceforth parenthetical citations to page numbers will be to this article): “we have or seem to have a clear positive understanding of the factual difference (or an apparent factual difference) between P1 and P2. If the future is such that P1 will be rendered true, then Andrea will wake up with the L-body, she will see the world from the L-bodies perspective: *she* will be the one who suffers if the L-body is damaged. But if P2 correctly describes what will happen, then Andrea will have quite different visual experiences when waking up (the ones connected with the R-body): she will act with the R-body, and she will live the life of the person who has the R-body.” Further (p. 196): “The difference appears to be factual in this sense: ‘D and Andrea is L-Andrea’ and ‘D and Andrea is R-Andrea’ are not just two legitimate description[s] of one and the same course of events. Rather, there is—according to the way we conceive of the situation—an objective possible feature of the world that makes one of the two descriptions true and the other wrong. The factual difference may be described [by] pointing out that Andrea will have a different future depending on which of the two possible identity facts will obtain.” I believe that NR has in mind Parfit’s claim that, on the Reductionist View, P1-P3 “are merely different descriptions of the same outcome” and the question whether one is one or the other or neither “is an empty question” because “it does not describe different possibilities, any of which might be true, and one of which must be true” (Parfit 1984, pp. 260–1). Parfit appears to agree with NR that admitting “different possibilities” in fission cases requires us to be “separately existing entities, such as Cartesian Egos” (p. 258).

(that I am him and he will have a headache settles that I will), but instead we infer identity on the basis of what we take to be correct future directed self-attributions (that it will be my headache tomorrow morning settles that I will be him).⁷

The second is that the content of future directed self-attributions of properties does not depend on our theories about the empirical criteria for transtemporal personal identity (e.g., material continuity versus psychological continuity). What scenario we are imagining remains the same through changes in our views about the empirical criteria. This allows us to change our minds about who we would be in various scenarios by fixing the facts relative to which we are considering the adequacy of different empirical criteria for its obtaining. It is the priority of future directed self-attributions to claims of transtemporal identity and the independence of their content from empirical criteria of transtemporal identity that underlies our recognition of the distinctions of the two situations expressed in P1 and P2, according to NR. We are, it seems, free to entertain the possibility that we are the subjects of the experiences of either.

These two points lead to a third, namely, that transtemporal self-identification is conceptually invariant with respect to changes in one's empirical criteria (such things as bodily or psychological continuity) for personal identity across time. One could change one's mind about this without changing the content of one's thought, though one might decide that what one thought was true was in fact false as a result.

According to NR, all of these points contrast with claims about transtemporal identity of non-conscious entities. For non-conscious entities, transtemporal identifications are prior to property attributions: we must first locate the thing as the same to talk about its properties

⁷ NR puts it this way (p. 198): “you understand what has to be the case for your utterance ‘I will be the L-person’ to be true on the basis of your understanding of what would render your self-attribution ‘I will have property P in the future moment *m*’ true.” The thought, I believe, is that the question whether I will be the F, where that picks out something I know to be a subject of experience, is settled by whether I will be the subject of F’s experiences, and there is no other more basic criterion. It seems true that if I think that I will be the subject of the headache experienced by the F tomorrow morning, then I must think of myself as the F. This is because in thinking that I will have property P tomorrow, whatever it is, I am thinking about myself. And the reason this does not seem beholden to any further check, I believe, is that when I think about myself in thinking this I think about myself directly and not by way of any description (see (Ludwig 1996) in this connection). But it seems to me also, by this same token, that I do not deploy any positive conception of my own nature. And so, though there is nothing in the thought that precludes it, there may yet be something about what it is I am thinking about in thinking about myself directly that makes it impossible for me to be the F. This point is connected with the final point in section 3.

and that requires applying criteria of transtemporal identity. If we change our view of the criteria, it changes the concept we deploy, and hence the content of such thoughts is not invariant with respect to empirical criteria for transtemporal identity. These differences are supposed to be traceable to the fact that we have first person thoughts about ourselves, but there is nothing corresponding in the case of nonconscious entities. Further, if we consider a single celled organism, *C*, which is by hypothesis nonconscious, which splits into two duplicates by fission, *L-C* and *R-C*, we do not, NR says, feel that there is a factual difference between the claims that *C* is identical with *L-C* and *C* is identical with *R-C*. We will return to these claims below when we turn to evaluating the argument.

The point about first-person attributions, NR claims, extends to third person attributions. As we take the truth of future directed self-attributions to determine the truth of claims about transtemporal personal identity claims, so in the case of others we take the truth of their first person future directed claims to determine the truth of claims about transtemporal personal identity claims. Even in the case of creatures without the conceptual resources to self-attribute, NR claims we can make sense of counterfactuals such as ‘If *x* were capable of thinking “I will be in pain”, then it would be true,’ and in terms of these we can ground claims of subject identity across time. This then gives us a fourth claim: transtemporal attribution of properties to other subjects of experience is conceptually prior to transtemporal identification and conceptually invariant with respect to changes in our empirical criteria for personal identity across time.⁸

NR uses these points to argue that the costs of giving up the factual difference between P1 and P2 is to give up thinking of oneself and others as subjects of experience, because to think of something as a subject of experience is *inter alia* to think of it as a thing for which attributions of properties (using, as she puts it, the conceptual resources of first person thought) is prior to the question of transtemporal identity, and if there is such a thing, one cannot but accept that there is a factual difference between P1 and P2. Thus, to deny the factual difference is to maintain that creatures who think of themselves as subjects of experience are subject to an unavoidable cognitive illusion—the illusion theory—one that is involved in every thought about transtemporal identity of subjects of experience, for if the difference is an illusion, there are, NR claims, no subjects of experience in the relevant sense at all.

⁸ I compress three claims (4–6) NR marks separately into one here (pp. 204–5).

Suppose that we reject the illusion theory, and so accept that there is a factual difference between P1 and P2. NR argues for subject-body dualism by arguing that a number of materialist candidates for what the subject of experience is cannot account for there being a factual difference between P1 and P2, and then generalizing from those cases. The three cases she treats are that the subject is identical with her body, that functionalism is true and that we are identical with certain functional systems with identity conditions distinct from bodies, and that the subject is constituted by her body rather than being identical to it, in the way we may wish to say that a statue is distinct from the clay of which it is made but is constituted by it, so that it is not an immaterial body in any sense.

The argument in each case has the same form, the differences represented here as clauses a-c in step 2:

1. There is a factual difference between the claims P1 and P2.
2. If subjects of experience were
 - a. identical with material bodies,
 - b. bodies with an appropriate functional organization,
 - c. constituted by material bodies,then there would be no factual difference between the claims P1 and P2.
3. Therefore, Andrea is not (a) identical to Andrea's body/(b) her body as a functional system/(c) constituted by her body.

The argument for 2a is that there are empirical criteria for transtemporal identity of material bodies, but given the symmetry of the case, there cannot be a factual difference between P1 and P2 if we identify subjects of experience with material bodies. The argument for 2b is that if we assume that persons are bodies with appropriate functional states, we cannot make sense of a factual difference between P1 and P2 because there is complete empirical symmetry between the relevant material successor bodies, and any account that identifies Andrea with L-Andrea as opposed to R-Andrea on a functionalist account would have to appeal to some relevant empirical difference between them. The argument for 2c is that the constitution view is committed to transtemporal identity conditions being spelled out in terms of empirical relations, but once again these are symmetrical between the cases

with respect to all the relations that the constitution theorist can appeal to.⁹

Although NR does not do so, the argument can be generalized as follows. For this purpose, let us say that x is distinct from any thing of type Y iff x is neither identical with anything of type Y nor constituted by anything of type Y .

1. There is a factual difference between the claims P1 and P2.
2. If subjects of experience were not distinct from anything that has empirical transtemporal identity conditions, then there would be no factual difference between the claims P1 and P2.
3. If subjects of experience were not distinct from anything of a type that entails that it is *inter alia* material, then it would have empirical conditions for transtemporal identity.
4. Therefore, Andrea is distinct from anything of a type that entails that it is *inter alia* material.

3. Three Interpretations of the Factual Difference

I will focus attention on premise 2 of this generalized argument.¹⁰ In evaluating this argument, it is crucial to get clear on what is meant by saying that there is or is not a factual difference between P1 and P2, repeated here.

P1: D and Andrea is L-Andrea.

P2: D and Andrea is R-Andrea.

⁹ More specifically, NR assumes that the proponent of the constitution view is committed to the following principle: "If B is the body of a person P at a given moment m and there are two human bodies $B1$ and $B2$ at a moment m' , and if $B1$ but not $B2$ constitutes the person P at m' , then $B1$ and $B2$ must be different with respect to their empirical relations to the body B that originally constituted person P " (p. 209).

¹⁰ It is worth noting that premise 3 can be sustained only if we add that if subjects of experience were not distinct from anything of a type that entails that it is *inter alia* material, it would be identical to something material that would be subject to fissioning. I do not see any reason to think this is *necessarily* so. The argument, however, might be reformulated so that the conclusion was limited to material things subject to fissioning. As the most plausible candidates for material things we are identical with are, it seems, in principle subject to fissioning, we would still have a result of significance if we accepted it.

One plausible way of reading ‘there is a factual difference between P1 and P2’ is as the claim that there is a *fact of the matter*, in the sense that one of P1 or P2 is correct. I do not think this is the interpretation that NR intends, but it is useful to consider it so that it can be clearly distinguished from other ways of understanding the claim and set aside. If this were the claim, then it would be difficult to sustain a materialist account of the subject of experience, because there could be no materialist ground for the transtemporal identity. In the present context, though, it is not easy to see how this could be persuasive because it seems at least as plausible to suppose that Andrea does not survive the operation, and, thus, it seems simply to beg the question against the ontic materialist. The fact that NR holds that P3

P3: Andrea is not L-Andrea and Andrea is not R-Andrea.

is a possibility as well as P1 and P2, and that she does not argue that this possibility is not actual, shows, I think, that this is not the interpretation she intends. Nonetheless, if one is minimally an ontological materialist, one must, I think, maintain that in the kind of case under consideration P3 is true.

A second, more plausible, way of reading the claim is as the claim that there is a difference in the content of the two claims, and this is the suggestion I made above. That there is a difference in content between the two claims, I believe, is correct, and it is compatible with P3. But on this way of reading it, it is unclear why it is not also true that there is a difference in the content of the two claims P4 and P5 (to just choose one of the candidates).

P4: D and L-Andrea body’s is Andrea’s body.

P5: D and R-Andrea’s body is Andrea’s body.

If there are in fact empirical criteria for transtemporal identification of bodies, then we have simply to interpret these claims in their light to given them content. And as the right hand conjuncts in P4 and P5 clearly differ in content, so do P4 and P5. If this is right, then just the fact that there is a difference in factual content between P1 and P2 could not show that Andrea is not identical with her body, and the point generalizes to any materialist proposal.

What could be the ground for saying that there is no factual difference, on the current view of what that comes to, between P4 and P5? Let us consider three possibilities with respect to how our criteria for transtemporal identity are fitted out to deal with such cases.

- 1) Our criteria for transtemporal identity for bodies tell us to endorse both P4 and P5.
- 2) Our criteria for transtemporal identity for bodies are silent on whether to endorse P4 or P5.
- 3) Our criteria for transtemporal identity, making provision for such symmetries, and sensitive to the fact that identity is an equivalence relation, tell us to reject both P4 and P5.

On the third, we reject P4 and P5 as both false, though differing in content, and there is no obvious ground for saying that they do not differ in factual content, though both are false. If either the first or the second of these were correct, though, there might be a ground for claiming that P4 and P5 do not differ in content.

On the first, (1), our criteria would be incoherent. We might then say that this means that there is no factual difference between them because ‘material body’ does not express a concept at all. But then there would be no factual content to any claims about material bodies in this case, which seems an unwelcome commitment. It would, in any case, surely be compatible with the materialist stance to revise our criteria, if they were found to be incoherent, so as to make more appropriate provisions for these sorts of cases. The simplest way to do that would be to stipulate that in cases in which there is an n -way tie ($n > 1$) for the best continuer, the original body does not survive.¹¹ *Perhaps* this will involve a revision of our notion of a material body (or a cell, etc.)—or at least a revision of the meaning of the term ‘body’ (etc.). But it will not change any verdicts delivered by the old usage, and it would cover the problem cases. And there would seem then no difficulty for the materialist to reinterpret ‘body’ throughout his discourse and then assert that P4 and P5, so interpreted, differ in content, and that they are both false in the envisioned circumstances.

¹¹ This is the line taken by Wiggins in his account of personal identity over time (Wiggins 1967, p. 55). Is it an absurd consequence of this view that in duplication cases whether the original subject survives depends on whether both duplicates rather than only one survive? What would we say in the case of cell fission? If we want a coherent account of cell identity over time, we cannot say the original cell survives as two. But then if we want to say it survives certain operations, which could have produced a duplicate (in the sense in question), we are committed to maintaining identity over time can be relative to whether or not something distinct survives. We can, of course, say that no cell survives any process that might have produced a duplicate, though it may be difficult to say what counts; but in any case we can say this in the case of persons as well. The bottom line is this: *whatever* we say about cell fission to resolve the parallel objections will work for a materialist account of personal identity over time.

This would be to modify the criteria so that they rendered the verdict in option (3). So far as the resulting view is a perfectly respectable materialist position, it would succeed in rebutting the charge that the materialist cannot make sense of there being a factual difference between P1 and P2, on this interpretation of ‘factual difference’.

On the second, (2), we imagine that the criteria give no verdict on the case. Then we might say that this means that there is no factual difference because, given our rules, nothing in the world determines that the second conjuncts of P4 and P5 are true or false. This gives some content to the idea that there would be no factual difference on this view. But this also shows that the practice with ‘material object’ is incomplete, and the sensible thing, as in the case of option (1), would be to complete the rules along the lines of (3). This would be a perfectly respectable materialist position, and so this would succeed in undercutting the force of the claim that P1 and P2 differ in factual content against materialism.

But this second reading may not what NR has in mind either. For there is still a possibility that seems open that would not be open on any materialist view, and which seems to be what informs NR’s discussion of what the factual difference comes to.¹² For on the maneuver we just considered, P4 and P5 are not possibilities at all. That is, on option (3), they are necessarily false. Yet, *prima facie*, P1 and P2 are possibilities. Thus, there is still available a sense in which P1 and P2 can be said to express a factual difference that P4 and P5 do not. P1 and P2 each express genuine possibilities, while P4 and P5 do not.

However, if this is the sense that we give to there being a factual difference between P1 and P2, the argument for subject-body dualism

¹² Parfit says that when “we ask an empty question, there is only one fact or outcome that we are considering” and that different answers are “merely different descriptions of this fact or outcome” (Parfit 1984, p. 214). Further: “When an empty question has no answer, we can decide to *give* it an answer” but that this is “not a decision between different views” (p. 214). Consider the question, with respect to the circumstances as described by ‘D’, whether Andrea’s body is identical with L-Andrea’s body or with R-Andrea’s body. I think Parfit would call this an empty question on the grounds that all the fundamental facts that can be appealed to to settle questions about transtemporal identity of bodies are already in view in what is expressed by ‘D’, irrespective of which of (1)-(3) obtains. Remaining questions are verbal, in the sense that they are questions about how to use words on the basis of the fundamental facts, for there is nothing else, in principle, to appeal to. This is what I think Parfit has in mind by saying different answers are descriptions of the same facts. To say they are verbal in this sense need not be to say it doesn’t matter how we answer the question, for I have argued that only option (3) makes sense for the materialist. But it is to say that the facts expressed by ‘D’ settle the matter. Now consider the question whether P1, P2 or P3 obtains. Is this question empty? If one says ‘yes’, then, given our interpretation of this, one is committed to saying that the facts expressed by ‘D’ are the only facts there are to appeal to in principle. This is to deny even the possibility that a subject of experience in such circumstances can be identical with one of the two subsequent persons.

appears to be unsound. The premise I wish to focus attention on, in the generalized argument, repeated here, is premise 2.

2. If subjects of experience were not distinct from anything that has empirical transtemporal identity conditions, then there would be no factual difference between the claims P1 and P2.

Let us rewrite 2 with the present interpretation of ‘no factual difference’ made explicit, as in 2’:

2’. If subjects of experience were not distinct from anything that has empirical transtemporal identity conditions, then the claims P1 and P2 would not be genuine possibilities.

In the heuristic of possible worlds talk, 2’ holds that if in the actual world the subject of experience were not distinct from anything that has empirical transtemporal identity conditions, then there is no possible world in which (P1) D and Andrea is L-Andrea or in which (P2) D and Andrea is R-Andrea. What would prompt us to accept 2’? If (N) were true,

(N) If P1 is possible or P2 is possible, then necessarily, if D, then Andrea is L-Andrea or Andrea is R-Andrea.

then 2’ would follow immediately, on the assumption that empirical conditions for transtemporal identity cannot sanction Andrea being identical to either L-Andrea or R-Andrea. For if, in any world in which D, Andrea were identical with something with empirical transtemporal identity conditions, then Andrea would not be identical to L-Andrea or to R-Andrea. The consequent of (N) would then be false, and from (N) we could conclude that neither P1 nor P2 is possible. But (N) is not true, or at least we have been given no reason to think it is. For there is at least one more possibility, namely, P3: D and Andrea is not L-Andrea and Andrea is not R-Andrea. In any world in which P1 or P2 obtains, if D, then Andrea is L-Andrea or Andrea is R-Andrea. If in every world P1 or P2 obtained, then (N) would be true. But we allow some worlds in which P3 obtains instead of either P1 or P2. In such a world, it is false that if D, then Andrea is L-Andrea or Andrea is R-Andrea. Hence, (N), if its antecedent is true, as we are supposing, is false, because it has a false consequent.

Suppose that in the actual world, then, D and Andrea is not L-Andrea and Andrea is not R-Andrea. We have no reason, so far, to think that in the actual world Andrea is not identical with or constituted by something that has empirical criteria for transtemporal identity. For Andrea not being either L-Andrea or R-Andrea is *prima facie* compossible with having empirical criteria for transtemporal identity

over time. Thus, it would appear as if there is a modal error in the argument, once we have characterized the content of the claim that there is a factual difference between P1 and P2 so as to provide the relevant contrast with objects that have transtemporal identity conditions. The error is to slip from the possibility of a world in which the subject of experience must be distinct from anything that has empirical transtemporal identity conditions to its actuality.

But not so fast! For are we not forgetting the necessity of identity (NI)?

(NI) For any x and any y , if $x = y$, then for any possible world w , if x exists in w or y exists in w , $x = y$ in w .

If we say that in the actual world Andrea exists and is not identical with L-Andrea or R-Andrea, then, given (NI), we have to deny that Andrea is identical with L-Andrea or R-Andrea in any possible world. Thus, we rule out, after all, the two possibilities that are introduced by the first premise of the argument.

This rejoinder, however, is altogether too powerful. For given that Andrea cannot be identical to both L-Andrea and R-Andrea, we must by this reasoning also deny that P1 and P2 are both possibilities. But there is no ground for admitting the one possibility without admitting the other. Thus, we should, it seems, admit neither. In that case, though, we get the conclusion that it is necessarily the case that Andrea is not identical with either L-Andrea or R-Andrea, and the case against materialism collapses.¹³

One could retreat to the claim that either it is possible that Andrea is L-Andrea or it is possible that Andrea is R-Andrea, but not both, and that which of these is possible is a brute fact. But then this equally requires denying that Andrea could fail to be whichever it is possible that she be, and so requires denying that it is possible that Andrea is neither. This would provide an immediate argument against the materialist, but it would have a different form than the argument above. More importantly, however, it is at least as *prima facie* plausible to think P3 is a genuine possibility as it is to think that either P1 or P2 is, and if we choose P3 over either of the others, then the materialist faces no challenge. So to complete this argument one would have to advance a reason to reject P3 as possible.

There is, perhaps, a way to rescue the argument from these difficulties. That is to see the possibilities as epistemic possibilities. How can

¹³ This requires thinking that if A at $t = B$ at t' , then for any world w , if A exists at any time at w or B exists at any time at w , then at w , at any times t_1, t_2 , A at $t_1 = B$ at t_2 . We assume here also an S5 modal logic, so that if it is possible that p , it is necessary that it is possible that p .

they both be epistemic possibilities if they are not both genuine possibilities? The idea is this: it is metaphysically or conceptually possible for subjects of experience to survive operations that result in fissions of the sort being entertained in the scenario involving Andrea. And it is metaphysically or conceptually possible for a subject of experience in a circumstance of the relevant type to be identical to the L-successor and likewise it is possible for a subject of experience in a circumstance of the relevant type to be identical to the R-successor. In either case, for such a subject to survive, it cannot be identical with any object with empirical criteria of transtemporal identity. If it is not, on the assumption that material objects are essentially material objects, it follows that necessarily it is not a material object. When one is thinking about P1 and P2, one is thinking of Andrea qua subject of experience. It is epistemically open that Andrea realizes an L-type survivor or an R-type survivor, or neither one. Whichever one it is, however, if either, fixes the facts about identity involving Andrea across all possible worlds.

This preserves the central idea that the conceptual structures associated with thoughts about ourselves allow that subjects of experience can survive fission as either of the successors, but allows that for any given subject of experience, whichever it is, if either, the subject is necessarily identical with it. On this view, we would restate premise 2' as in 2''.

2''. If subjects of experience were not distinct from anything that has empirical transtemporal identity conditions, then the claims P1 and P2 would not be genuine epistemic possibilities.

Unfortunately for the argument, however, 2'' is false. For it is clearly compatible with Andrea being identical with or constituted by a material thing that it is *epistemically* open for her (and for us) that she is not. We are back to a form of Descartes's argument from doubt.¹⁴ We have, then, found to no reading of 'factual difference' on which the argument is successful.

4. Ontic versus Reductive Materialism

We have been granting that it is possible for a subject of experience to survive as one of the candidates in a fission operation, as well as to fail to survive. This generates three epistemic possibilities when it is not settled for us what our fundamental natures actually are. The trouble with the argument is that one of the epistemic possibilities is compatible with our being material things. Thus, the argument simply leaves open

¹⁴ See (Malcolm 1965). Although sometimes attributed to Descartes, it is unclear whether Descartes was committed to any such argument (Cottingham 1986, 112–3).

whether as a matter of fact we are or are not identical with material things.

But if we grant the underlying idea, there is still a conclusion that can be drawn that is significant. It is implicit in what we have said already. Suppose that a version of reductive materialism is true, that is, that mental states, and conscious states in particular, are reducible to either physical states or functional states or perhaps function-cum-relational states. Then subjects of experience will of necessity have empirical criteria for transtemporal identity. And if this is right, then it cannot be that it is possible that subjects of experience survive fissioning. Thus, if it is possible, reductive materialism is false. The argument goes as follows:

1. Necessarily, if reductive materialism is true, then necessarily mental properties would be first-order material properties of objects or second-order functional-cum-relation properties of objects.
2. Hence, necessarily, if reductive materialism is true, necessarily, subjects of experience would be material objects or functional systems.
3. No material object or functional system can survive fissioning.
4. Subjects of experience can survive fissioning.
5. Therefore, reductive materialism is false.

The apparent intelligibility of survival in fission thought experiments does therefore present a serious challenge, not to ontic materialism, but to reductive materialism. The materialist either must reject the possibility of survival for a subject of experience, or reject the view that material objects have purely empirical criteria for transtemporal identity. If NR is right, then it is not possible to deploy the concept of a subject of experience without seeing it as a possibility. Granting this, a modified form of NR's conclusion is still available: to be a reductive materialist and embrace empirical criteria of transtemporal identity, one must conclude that we are subject to an ineluctable illusion in so far as we conceive of ourselves as subjects of experience.

References

- Baker, Lynne Rudder. 2000. *Persons and bodies: a constitution view, Cambridge studies in philosophy*. Cambridge: Cambridge University Press.

- Cottingham, John. 1986. *Descartes*. Oxford: Blackwell.
- Hasker, William. 1999. *The emergent self*. Ithaca, N.Y.: Cornell University Press.
- Lowe, E. J. 1996. *Subjects of experience*. New York: Cambridge University Press.
- 2006. Non-Cartesian Substance Dualism and the Problem of Mental Causation. *Erkenntnis* 65:5–23.
- 2010a. Substance Dualism: A Non-Cartesian Approach. In *The waning of materialism*, edited by G. Bealer and R. Koons. Oxford: Oxford University Press.
- 2010b. Why My Body is Not Me: The Unity Argument for Emergentist Self-Body Dualism. In *Emergence in science and philosophy*, edited by A. Corradini and T. O'Connor. London: Routledge.
- Ludwig, Kirk. 1996. Singular Thought and the Cartesian Theory of Mind. *Noûs* 30 (4):434–460.
- Malcolm, Norman. 1965. Descartes's Proof that his Essence is Thinking. *The Philosophical Review* 74 (3):315–338.
- Nida-Rümelin, Martine. 2007. Dualist Emergentism. In *Contemporary debates in philosophy of mind*, edited by B. McLaughlin and J. Cohen. Oxford: Blackwell.
- 2010a. An Argument from Transtemporal Identity for Subject-Body Dualism. In *The waning of materialism*, edited by G. Bealer and R. Koons. Oxford: Oxford University Press.
- 2010b. What About the Emergence of Consciousness Deserves Puzzlement? In *Emergence in Science and Philosophy*, edited by A. Corradini and T. O'Connor. London: Routledge.
- 2006. Doings and Subject Causation. *Erkenntnis* 67 (2):147–372.
- O'Connor, Timothy and Jonathan D. Jacobs 2003. Emergent Individuals. *The Philosophical Quarterly* 53 (213):540–555.
- Parfit, Derek. 1984. *Reasons and persons*. Oxford: Clarendon Press.
- Spinoza, Benedict de. 1994. *A Spinoza reader: The ethics and other works*. Translated by E. Curley. Princeton: Princeton University Press.
- Swinburne, Richard. 1984. Personal Identity: The Dualist Theory. In *Personal identity: Great debates in philosophy*, edited by S. Shoemaker and R. Swinburne. Oxford: Blackwell.
- 1986. *The evolution of the soul*. Oxford: Oxford University Press.
- Wiggins, David. 1967. *Identity and spatio-temporal continuity*. Oxford: Oxford University Press.
- Zimmerman, Dean and Penelope, Mackie. 2010. Mind-Body Dualism. *Proceedings of the Aristotelian Society* LXXXIV:119–149.