

Are We Causally Redundant? Eliminativism and the no-Self View

JIRI BENOVSKY
*Department of Philosophy,
University of Fribourg, Fribourg, Switzerland*

Some friends of eliminativism about ordinary material objects such as tables or statues think that we need to make exceptions. In this article, I am interested in Trenton Merricks' claim that we need to make an exception for us, conscious beings, and that we are something over and above simples arranged in suitable ways, unlike tables or statues. I resist this need for making an exception, using the resources of four-dimensionalism.

Keywords: Eliminativism, material objects, Trenton Merricks, consciousness, causality.

The eliminativist view about ordinary macroscopic objects like chairs or statues suits the taste of those who prefer desert landscapes to baroque complexity, and it nicely solves a number of problems (composition, vagueness, material constitution, coincidence, causal overdetermination,...).¹ It elegantly avoids these problems with ordinary objects since if there are no such objects, there are no worries concerning them. The issue I will be interested in this article is to see whether eliminativists need to make an exception—for us. Indeed, eliminativists such as Peter Van Inwagen or Trenton Merricks famously argued (for different reasons and in different ways) that while eliminativism is the best theory around when it comes to tables, planets, or statues, it is not to be endorsed in the case of humans—an exception is to be made. Van Inwagen focuses on *living* entities, and Merricks focuses on *conscious* organisms. In this article, I will examine Merricks' reason to make such an exception, and I will argue that it can be resisted.

The question is: can we eliminate the Self in the same—or similar—way we can eliminate tables and statues, without losing some-

¹ See Unger (1979), Van Inwagen (1990), Merricks (2001), and Heller (1990, 2008).

thing important? Under ordinary-objects eliminativism, we can without any loss eliminate tables and statues because there are simples arranged tablewise or statuewise and those can play the same practical and theoretical roles that tables or statues could play if they existed. (Eliminativism is also compatible with ontologies that do not postulate the existence of simples, but I will leave this issue aside here; see Benovsky (2016) for a detailed discussion.) In this view, chairs understood as *single* objects, can be eliminated because there is a *plurality* of objects that takes their place, namely, simples arranged chairwise. I believe that the same strategy can be applied to the case of the Self, although it needs to be articulated in a way that suits such a special case. The basic idea is the same: a single entity such as *the* Self can be eliminated because there exists a plurality of other entities, namely, successive impermanent psychological states/experiences arranged ‘Selfwise’. I have articulated in detail and defended this view in Benovsky (manuscript); here, my aim is to defend it against an objection raised by Merricks (2001).

The idea of such an eliminativism about the Self is that we can be eliminativists about *us*, understood in a reified and ontologically committing sense of Selves, but that we don’t lose anything—I can still say “I am drinking a beer “ in a sense understood in terms of simples arranged my-body-wise and beerwise, and in terms of the existence of a succession of impermanent psychological states. This is how we can hold a unified and complete eliminativist view, with no exceptions. In Merricks’ (2001) view, however, there is a *disanalogy* since entities like tables or statues are *causally irrelevant*—whatever they can cause, can be caused entirely by the simples that compose them. So, in his line of thought, this is one of the good reasons to say that tables do *not* exist. But, he adds that “we humans—in virtue of causing things by having conscious mental properties—are causally non-redundant” (Merricks 2001: 114). When I decide to run, there is a cause to be understood in terms of microphysics or microbiology, but there *also* is a cause to be understood in terms of my *decision*. It is my decision, Merricks says, that causes the simples to move as they do. This is why we cannot be eliminated in the same way tables can be, since we are causally relevant—we have causal powers over and above the causal powers of simples that compose us. Merricks’ argument to the effect that we, human organisms, are causally non-redundant in virtue of having conscious mental properties is a complex and a very long one—indeed, it stretches on almost thirty pages (see Merricks 2001: Chap. 4). In what follows, let me focus on a (rather self-standing) part of his argument in detail and see how this step can be resisted—if it can, the overall argument will then not go through.

The main point of the argument is to show that we humans are not causally redundant, and that we have conscious mental properties that do “not supervene on what our parts are like” (Merricks 2001: 88). We

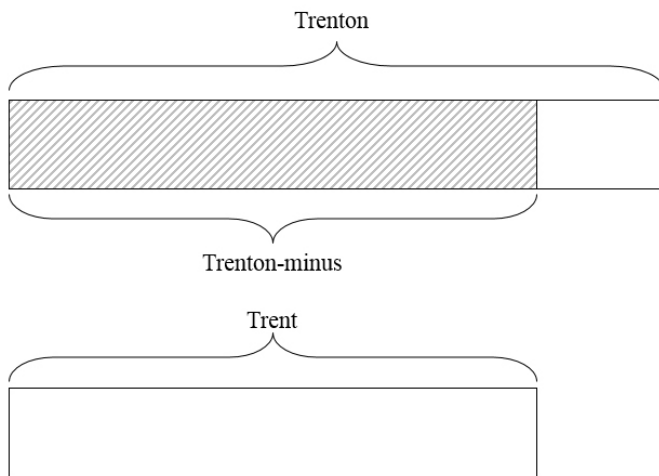
cause things in virtue of having these properties and we are therefore not causally redundant. (Thus, an argument based on the idea that we should eliminate anything that is causally redundant cannot be used to eliminate us—this is Merrick’s overall main point.) On the way of defending this claim, Merricks argues for the rejection of:

Consciousness (C). Necessarily, if some atoms $A_1 \dots A_n$ compose a conscious object, then any atoms intrinsically like $A_1 \dots A_n$, interrelated by all the same spatiotemporal and causal interrelations as $A_1 \dots A_n$ compose a conscious object. (Merricks 2001: 94)

Here is, in short, Merricks’s argument against (C)—it is a variant of the ‘undetached parts argument’.² Suppose a small part of you is annihilated (say that your finger, or perhaps just an atom composing your finger, is cut away). Right after the amputation there is a conscious object—you—composed of some atoms. But these atoms existed in exactly the same way just before the amputation—indeed, they composed a (big majority) part of you. But, if (C) is true, this means that even before the amputation there was an object that was part of you—let’s call it “you-minus”—that was a conscious object. So, it seems that before the amputation there were *two* non-identical conscious objects, namely you *and* you-minus. (By the same reasoning, there actually were *many* you-minus-like objects before the time of the amputation.) But this is false, since there was only one conscious object before the amputation. Thus, Merricks concludes, by *reductio* (C) is false.

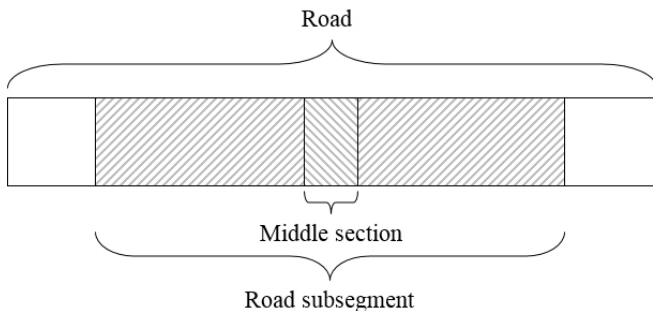
In case one would be tempted to answer the objection by appealing to four-dimensionalism and using talk about temporal parts to escape the unwelcome consequence that there were two conscious objects before the amputation, Merricks provides a temporal version of the objection as well, which can be formulated as follows. Take a four-dimensional person named “Trenton” who lives for 100 years. Take also another four-dimensional person, inhabiting the same possible world, who lives for only 80 years and is named “Trent”. Suppose that Trent is microphysically intrinsically exactly like the temporal part of Trenton who lives for the first 80 years of Trenton’s existence, and let us call this temporal part of Trenton “Trenton-minus”, where Trenton-minus and Trent thus have atomic temporal parts exactly similar in intrinsic features and causal and spatiotemporal interrelations.

² See Van Inwagen (1981); for a discussion see *inter alia* Heller (1990) and Benovsky (2006).



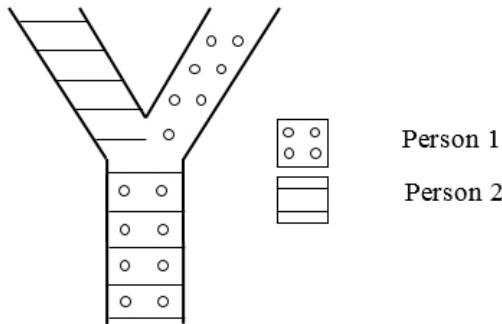
According to Merricks, according to the four-dimensionalist, assuming (C) for *reductio*, when it comes to Trenton between the age of 0 and 80, we then have a case where there are *two* coincident persons: Trenton and Trenton-minus. Trenton-minus is a conscious person in virtue of the existence of Trent and in virtue of the truth of (C). But such coincidence is unacceptable, and as a consequence, by *reductio*, (C) is false.

But the way Merricks presents the case here can be resisted. Indeed, this is not how four-dimensionalists typically describe the situation. Here is Sider (2001: 6), about the Statue and Lump famous case of coincidence: “At any given time it is only a temporal part of a spacetime worm that is wholly present. Thus it is only temporal parts of Statue and Lump that are wholly present at the time of coincidence. How can these temporal parts both fit into a single region of space? Because ‘they’ are *identical*.” Sider then compares this to a case of a road:



There is a road that has a subsegment. In the very middle of the road, should we say that there are *two* entities, namely, the road and the road subsegment? Of course not. There is only one entity—the middle section—common to both the road and the road segment.

Similarly for Trenton and Trenton-minus. At a time where ‘both’ Trenton and Trenton-minus exist, should we say that they are two persons? Of course not. Before Trenton’s 80th birthday, there always was only one person, exactly as in the middle section of the road there is only one road. It’s just that this middle section is part of a road and of a road subsegment, and in the same way Trenton-minus is part of Trenton. This does not prevent Trenton-minus to be a person, as for instance David Lewis insists upon: “A person-stage is a physical object, just as a person is. (If persons had a ghostly part as well, so would person-stages.)” Lewis (1983: Postscript B). Typical examples of the way four-dimensionalism deals with such situations also involve cases of fission.³ Let us say, for the sake of brevity, that for some reason a person undergoes fission, perhaps using a transporting device such as the one commonly used on the *USS Enterprise*, where due to a malfunction of the device, instead of simply transporting one person from one place to another, the device *also* leaves the original person behind. (You can replace this example with any other case of fission, if you don’t like Star Trek stories.) Thus, after the fission, there are two persons, exactly alike. According to four-dimensionalism, we then have a situation where there are two four-dimensional persons, sharing an initial segment:



³ I discuss one such case in detail in Benovsky (2013: 162–164).

Under a typical *endurantist* reading, this situation is one where the threat of coincidence is real: Person1-after-fission is not identical to Person2-after-fission, but Person1-after-fission is identical to Person1-before-fission, and Person2-after-fission is identical to Person2-before-fission—we then seem to have a situation where Person1-before-fission *is* identical to Person2-before-fission, that is, where these two persons seem to coincide in an unpalatable sense. But not so under the four-dimensionalist view where Person1-after-fission is *not* identical to Person1-before-fission, since these are two different temporal parts, numerically distinct, and similarly for Person2. In this way, (i) four-dimensionalists do not have to face the threat of coincident entities,⁴ and (ii) they can say, relevantly to our present discussion, that there is only one person before the fission, in the same sense that there is only one person in the case of Trenton and Trenton-minus, and similarly in the spatial case of you and you-minus (in the finger amputation case). Metaphysically speaking, in all such cases where there seem to be two objects competing for the same space (and time), there really is only one, it's just that it's also part of other, spatially and/or temporally bigger, objects. It's like a wall that's common to two houses: if you need to repair it, you'll only need bricks to repair *one* wall, not two. In the way Merricks describes the situations he uses in his argument, there seems to be something like the principle that only 'the biggest' object is the one that counts. Thus, only Trenton, but not Trenton-minus is a conscious object. In his argument, appealing to the existence of Trent and to the truth of (C), Merricks then wants to force his opponent to recognize, for *reductio*, that Trenton-minus *is* a conscious object as well, thus creating a situation where there apparently are two distinct coincident objects that crowd each other out. But, as we have seen, four-dimensionalists do not, and do not have to, understand this situation (as well as the other similar situations) in this way.

One issue still remains. Who is doing the thinking, in the four-dimensionalist view? Is it the whole worm, or is it only a (rather short-lived) temporal part of the worm? If it were the whole worm, then—and only then—Merrick's objection above would go through. So, in order for the reply to work, we have to say that it is not the whole worm that has thoughts but that it has them only in virtue of having temporal parts that have them. A possible objection arises here:⁵ say that a temporal part that lasts for only one minute thinks the thought "I have lived for 50 years". This is true, in a sense, because the whole worm did live for 50 years. But it is false, when thought by the temporal part, since the temporal part only lived for one minute. So, the same thought seems to be both true and false—how messy!

⁴ I simply assume here, in agreement with Merricks, that such coincident entities are not acceptable.

⁵ I would like to thank Trenton Merricks for raising this point in a discussion.

But this only points to a specific feature of four-dimensionalism (i.e. the worm view⁶). In this view, worms have most of their properties in virtue of the having of those properties by their temporal parts. The temporal parts can overlap, in the sense we have just seen, or in a fission scenario. Thus, what is being thought or said at some point by some temporal part is *ambiguous*. Take the case of fission we have seen above. Let us say that, *before* the fission, the person that is there is called “Jean-Luc Picard”. From an atemporal standpoint such a name is then ambiguous, since it refers both to Person1 and Person2, and since these two overlap at the time before fission, but not at the time after the fission. But as David Lewis points out, such an ambiguity is perfectly harmless as long as the two bearers of the name “Jean-Luc Picard” are indiscernible—that is, precisely, before the fission (see Lewis 1983: 64–65). The need to distinguish the two persons arises only *after* the fission, and there is no ambiguity there, since there clearly are two persons, and we will use two different names to refer to them (even perhaps in a homonymic way). Similarly, what the one-minute-long temporal part thinks or says is ambiguous, and it is true under one disambiguation and false under another. The problem then easily dissolves as a mere case of ambiguity.

As a consequence, using four-dimensionalism to answer Merricks’s objection, we can say that (C) is *true*, and thus one cannot use the alleged falsity of (C) to argue to the effect that we humans are not causally redundant because our conscious mental properties do not supervene on what our parts are like. And one cannot then use this as a reason to make an exception for us when it comes to eliminativism.⁷

References

- Benovsky, J. 2006. “Four-dimensionalism and modal perdurants.” In P. Valore (ed.), *Topics on General and Formal Ontology*. Polimetrica Publisher.
- Benovsky, J. 2013. “Branching and (in)determinism.” *Philosophical Papers* 42 (2): 151–173.
- Benovsky, J. 2016. “Eliminativism and gunk.” *Teorema* 35 (1): 59–66.
- Benovsky, J. manuscript (draft). “Subjectivity and the no-Self view: How to survive eliminativism about the Self.”

⁶ In this article, I have focused on four-dimensionalism understood as the view that objects like tables or people are temporally extended entities, composed of temporal parts, since this is the view that Merricks appeals to in his argument, and since it is, to my mind, the best available version of four-dimensionalism. But note that another variant of four-dimensionalism, namely *the stage view* (see Sider (2001) and Varzi (2003)), can accommodate the truth of (C) even more easily, since according to this view conscious objects like persons are only instantaneous entities (that persist by having temporal counterparts at different times), and there is then no worry concerning amputation or fission cases since the person after the amputation or after a fission simply is a different person.

⁷ I would like to thank Baptiste Le Bihan, Damiano Costa, Mark Heller, and Trenton Merricks for very helpful comments on an earlier version of this article.

- Heller, M. 1990. *The ontology of physical objects: four-dimensional hunks of matter*. Cambridge: Cambridge University Press.
- Heller, M. 2008. "The Donkey Problem." *Philosophical Studies* 140 (1): 83–101.
- Lewis, D. 1983. "Survival and Identity." In Lewis, D. *Philosophical Papers*, vol. I. Oxford: Oxford University Press.
- Merricks, T. 2001. *Objects and Persons*. Oxford: Oxford University Press.
- Sider, T. 2001. *Four-dimensionalism*. Oxford: Clarendon Press.
- Unger, P. 1979. "There are no ordinary things." *Synthese* 41 (2): 117–154.
- Van Inwagen, P. 1981. "The doctrine of arbitrary undetached parts." In Van Inwagen, P. 2001. *Ontology, Identity and Modality*. Cambridge: Cambridge University Press.
- Van Inwagen, P. 1990. *Material Beings*. Ithaca: Cornell University Press.
- Varzi, A. 2003. "Naming the stages." *Dialectica* 57: 387–412.