

JOHN SCHWENKLER

NO WORK FOR A THEORY OF PERSONAL IDENTITY*

A main element in Richard Swinburne's (2019) argument for substance dualism concerns the conditions of a person's continued existence over time. Simplifying greatly, Swinburne's argument is that though there is always a fact of the matter as to whether a person who exists at one time is or is not the same person as a person who exists at some other time, all complex (physical, psychological, and mixed) theories of personhood are such that, concerning various real or imagined cases, they cannot yield a decisive verdict as to whether this identity relation holds. From this we are to conclude, first, that these complex theories are false, and, further, that since substance dualism does yield decisive verdicts about these cases, it is to this degree supported.

In this brief commentary I aim to question two things: first, whether the kind of imaginary cases that Swinburne relies on to make his case should be accorded the kind of weight he supposes; and second, whether philosophers should be concerned to give any substantial *theory*, of the sort that dualism and its competitors are apparently meant to provide, to explain the conditions of personal identity after all. My suggestion, instead, will be that the concept of a person's continued existence is better taken as philosophically unanalyzable.

JOHN SCHWENKLER, Professor of Philosophy at Florida State University; address for correspondence: Department of Philosophy, 151 Dodd Hall, FSU, Tallahassee, FL 32306; e-mail: jlschwenkler@gmail.com.

* I thank Ted Poston especially and also Taylor Carman, Tom Cochrane, Jonathan Folstein, Mary Gormally, Joseph Jedwab, Adam Omelianchuk, Gualtiero Piccinini, Brandon Towl, Jasper van der Herik, and Eric Winsberg for feedback and discussion

I

Let a *complex theory of personal identity* be a theory of personal identity which tries “to analyze the relation of personal identity, P_1 being the same person as P_2 , in terms of relations between various other features of P_1 and P_2 .”¹ Depending on the theory these features may be physical, psychological, or a mix of the two: for example, one might try analyzing personal identity as a matter of having the same bodily organs (most likely, the same brain), of having the same apparent memories or character traits, or of having the right combination of both of these kinds of things. And Swinburne’s argument against these theories turns on various thought experiments concerning what will happen if parts of a person’s brain, and with it some portion of her psychological characteristics, are either removed from her body or transplanted into the body of someone else. For example (I will quote at length because the details of the rhetoric are important to my criticism):

While a body may be said to be the same body when a kidney or heart is replaced, it is not clear whether we should call the body the same if many of its parts are replaced very quickly. And there is good reason to believe that every bodily part will soon prove replaceable. So, what if all the bodily organs are all replaced with the exception of the brain? Surely the resulting body would not then be the same body as before. To replace all the parts apart from the brain in one operation could be achieved most easily simply by transplanting the brain into the skull of some other body from which its brain had been removed. Maybe this would only be possible if the brain being replaced was very similar in many aspects of its structure to the brain being transplanted, as would be the brain of an identical twin. Surgeons are only just learning to reconnect severed brain nerves. Even when they have learnt to do this, it will be a lot longer before they learn to connect enough nerves to reconnect a whole brain to a new body—since a brain has so many nerves connecting it to the rest of the body. But there seems no reason in principle why this should not be achieved. Yet if surgeons take your brain and put it into another body, would the resulting body now be your body, or would it still be the body of the person whose brain had been removed? It would surely now be your body, since you could now move its parts in the same way as you could previously move the parts of your earlier body, and you would now learn about the world through the eyes and ears of the new body. And so the resulting person would surely be you.... So, it would seem, you go where your brain goes. (45–46)

Scenarios of the same kind are used to argue against complex theories that appeal to psychological characteristics or a mix of psychological characteristics and physical ones.

¹ Richard SWINBURNE, *Are We Bodies or Souls?* (Oxford: OUP, 2019), 42; my parenthetical page references are to this volume.

In particular, two crucial variants on the imagined process of “simply” transplanting an entire brain involve either moving just a part of a brain, and with it just a part of its original owner’s psychology, into a new body, or moving one part of brain into one body and the other part of it into another—with, it is assumed, the original owner’s psychology then being divided between the two resulting persons. The idea in each case is that complex theories of personal identity can’t give a reasonable account of the ensuing facts.

There are two things I want to say about this form of argument. The first is that I don’t think we do have much reason at all to believe what Swinburne says we have good reason to believe, viz., that it is possible to replace any and every part in a person’s body without killing them. And while I agree that it *might* be possible in principle to move one person’s brain into the body of another person who’s sufficiently physically similar to them, nor do I think we have much positive reason to think that this *is* even an in-principle possibility. Without these assumptions, the argument from transplant cases against complex theories of personal identity fails: the most we can say is that *if* these things are possible, *then* (perhaps—I will say more here below) those theories are not true. But we don’t have sufficient evidence to support that crucial assumption.

Wait—but what about all the observed cases of successful organ transplantation in humans? And isn’t the brain just an organ like any other, able to be kept alive for some time while separated from a body and then, perhaps, transplanted into another? The latter assumption is the one that I find unjustified, and I suspect that the inevitable attraction of it (which I also feel) is to a large degree the product of our already having accepted the picture of human nature that, for Swinburne, the possibility of total brain transplantation is supposed to help support. (This concern holds in spades for judgments about the supposed possibility of transplanting a brain only in part, or of transplanting one part of a brain into one body and another part into another.) At least to my knowledge, given the state of present science there is a serious possibility that the boundaries of “the brain” are not nearly as neatly defined as arguments of this kind require them to be, as brain structures may blend so seamlessly into other structures of the central nervous system, and from there into the peripheral nervous system as well, that transplanting “the brain” of one person into another’s body would not be, as Swinburne supposes, a matter of taking out what is in her head and then attaching the severed nerves to some different ones, but rather of taking out the brain *and* the attached nerves *and*, perhaps, many of the sensory and motor structures to which they are further connected.² The more

² For more detailed argument in this vein, with close attention to the scientific details, see Evan THOMPSON and Diego COSMELLI, “Brain in a Vat or Body in a World? Brainbound Versus Enactive Views of Experience,” *Philosophical Topics* 39, no. 1 (2011).

of this that is required to perform the imagined operation in a way that leaves us with a functioning human being on the other side, the more that transplanting “the brain” of the original person starts to look like it is just a matter of giving that person a new skeleton, or perhaps a new skeleton plus some new muscles and sensory and vital organs. And the further we go down this route the more that the description “take [a person’s] brain and put it into another body” starts to appear inapt.

I don’t mean at all to say that we *know* that the picture I’ve so very roughly sketched is the correct one, or even that we have good reason to think it truer to the biological facts than Swinburne’s alternative. The point is only that, at least by my lights, we don’t know that this picture *isn’t* correct, either—for the evidence is not yet in—and that absent this knowledge we don’t know that the possibilities on which Swinburne’s argument is premised are real ones. As I suggested above, the most we can say is that *if* the facts of human biology are as he supposes, then those facts raise trouble for these complex theories of personal identity. But it takes further argument, which Swinburne hasn’t supplied, to say whether the antecedent to this conditional is true.

Let’s set this objection aside, and concern ourselves only with the conditional, or perhaps suppose *arguendo* that, in the phrase of Philippa Foot, “philosophers have arranged” that it is possible to transplant a person’s brain in just the way that Swinburne imagines. My second objection concerns what Swinburne says would “surely now” be the case, assuming that such a thing were to transpire. *Can* we reasonably be so sure about what would happen in such a case, especially given that, as I’ve just emphasized, it is so far from the kinds of cases we have encountered in the past? I don’t think we can be. That is, I don’t think we can reasonably be sure *either* that, in the wake of brain transplantation, the person with the transplanted brain would be psychologically similar to the person whose brain it used to be, *or* that, assuming the resultant person was so similar, this would show that she *was* the person who had had this brain before the process of transplantation, rather than that she *took herself to be* this person, and was of course very psychologically like her. My doubt about the first point stems from the sorts of concern I expressed just above, concerning the remoteness of this (supposed) possibility from our direct experience, and so our ignorance about what it would really involve. There are a great many things about the structure of the physical universe, including the human body, that have totally confounded our prior expectations. Couldn’t it also turn out, contrary to our expectations, that when a brain is transplanted into a new body the psychological characteristics of its original owner don’t travel with it, but rather are erased, or changed very radically, during its removal? I don’t think there is anything we *know* about the brain that shows us that this wouldn’t happen. And concerning the second point my doubt stems from the kinds of arguments Bernard

Williams provides in his great paper “The Self and the Future”: it seems to me at least *conceivable* that the person who awoke from such an operation, thinking for instance how glad she was that she had survived the operation, and that she now had this new body to support her, would be entirely deluded in these judgments, as *she* had not existed for more than a few hours, or in any body other than this one. (It’s worth noting that on Swinburne’s own view judgments of personal identity don’t extend infallibly to the past: see pp. 107–8.) If that were the harsh reality then it might be impossible for anyone ever to tell that this was so—but then, it seems to me, the same point will hold if things turn out in the way that Swinburne imagines. For me, the upshot of these arguments is not any confidence that Swinburne’s assumptions are wrong, but rather a total lack of confidence that we are in any good position to say whether they are.

II

Complex theories of personal identity assume that we have a conceptual handle on one thing, say what it is for a body to remain the same body or for one person to have certain of the same psychological states as another, and then can use this concept to illuminate the conditions of a person’s identity over time. Let a *simple theory of personal identity* be a theory that denies this assumption. Consider, for example, the following Very Simple Theory of the conditions of a person’s identity over time:

(VST) P_2 is the same person as P_1 iff P_2 either remembers or has forgotten P_1 ’s past deeds.

This Very Simple Theory is not totally unilluminating—it is not, for example, the kind of theory we could give of the identity conditions of a table or a chair, since these entities aren’t such as to do things, let alone to remember their deeds or forget them. Yet it’s clearly not illuminating in anything like the way that the following Crude but Complex Theory would be, if the latter theory were true:

(CCT) P_2 is the same person as P_1 iff P_2 has most of P_1 ’s apparent memories.

I find it obvious that CCT is mistaken: it fails in the left-to-right direction since it’s possible to have total amnesia, and it fails in the right-to-left direction since it’s possible for a person to be *deluded* about her past, thinking herself to remember doing things that were done by a different person entirely. But the important point for present purposes is that CCT, unlike VST, is the sort of theory that *at least could* be

mistaken, since the concept “has most of X’s apparent memories” doesn’t *presuppose* the identity of the person it applies to with X, in the way that the concept “either remembers or has forgotten X’s past experiences” clearly does—since *only X* could either remember or forget those, and so to say of *someone* that she either remembers or has forgotten X’s past experiences is already to say that that person is X. And the question I want to consider now is whether there is good reason to demand that a philosophical account of personal identity avoid trafficking in concepts of this latter kind, or at least not traffic in them exclusively. To ask this is to ask: What reason is there for a philosophical account of personal identity to be simple rather than complex?³

Swinburne’s answer to this question is unfortunately a bit confusing, since over the course of his discussion of personal identity the working understanding of simple theories shifts a bit—from the canonical definition at the start of chapter 3 that I quoted earlier, according to which simple theories are those that try to analyze personal identity “in terms of relations between various other features” of the persons under consideration (42), to the definition at the end of that chapter of a “simple theory of personal identity” as one on which “there are no necessary or sufficient conditions for personal identity *in terms of the degree of any feature or which there can be degrees*” (65, emphasis added), to the even more restrictive claim at the start of chapter 4 that according to a simple theory “that a person P_2 [is] the same person as a person P_1 at an earlier time is not analyzable in terms of their having some of the same body or brain or other bodily parts, or in some kind of continuity between their physical or mental properties” (69). The shift seems to me to matter, since if the first definition is the correct one then Swinburne’s dualist theory of personal is *not* a simple theory, at least on the assumption, which I wouldn’t expect him to challenge, that a person’s soul is a “feature” of that person. In any case, his theory differs this much from one like the Very Simple Theory that I sketched just above, since as Swinburne understands it the concept “has X’s soul” is like “has most of X’s apparent memories,” and unlike “either remembers or has forgotten X’s past experiences,” in that to apply it to someone is not *already* to say that this person is X, and so the judgment that a person has, or doesn’t have, some person’s immaterial soul can be a way to *explain* how it is that they are, or are not, that person.

At the start of chapter 4 in his book Swinburne offers an argument that a theory like my Very Simple Theory, even if what it says is true, cannot be used to support a non-dualist position. The argument turns on a case where half of the brain of one

³ For related discussion, see John McDOWELL’s “Reductionism and the First Person,” in *Reading Parfit*, ed. Jonathan Dancy (Oxford: Blackwell, 1997).

person, who we may call “Alexandra,” is transplanted into the body of another person half of whose brain has been removed. Swinburne writes that were this to happen,

it would be logically possible that the resulting person is Alexandra and also logically possible that the resulting person is not Alexandra. So perhaps personal identity is totally unanalysable; either the resulting person is or she is not Alexandra, and that is all that can be said. That does not follow. If the resulting person is Alexandra, there must be some part of her which makes her Alexandra, which a resulting person who is not Alexandra would not have. (69)

There are a few points at which a defender of my Very Simple Theory could try to resist this framing of the situation. According to that theory we *can* say of the resulting person that she is Alexandra if and only if she remembers or has forgotten Alexandra’s past experiences, and that either the memory of these past experiences, or (perhaps) the lack of them, *is* in some sense a “part of her” which makes her the person she is, and which couldn’t be had by anyone else. But let’s set this aside, since as I have said the concept “either remembers or has forgotten Alexandra’s past experiences” is one that simply builds in a concept of personal identity, and so the “part” of the resulting person that the theory invites us to appeal to in saying whether that person is Alexandra or not isn’t one that does much to *explain* what makes her the person she is, let alone to *analyze* what personal identity consists in. Is there, though, any good reason to demand something more than this?

Here, as well as I can make it out, is Swinburne’s (68–70) argument that there is such a reason:

- (1) Humans are composite substances.
- (2) So the ‘principle of the identity of composites’ applies to humans: at any time any human with a certain body, made of certain component parts, arranged in the same way, each of the parts having a certain past history, and certain essential and non-essential mental and physical properties, would be the same person.
- (3) In the thought experiment described above, the resulting person is either Alexandra or some human who is not Alexandra.
- (4) Whichever person she is, there must be something about her which distinguishes her from the other.
- (5) Yet this distinguishing element cannot be a physical part of that person, since the resulting person will have the same component physical parts, arranged in the same way, each with the same past history, whether she is Alexandra or some human who is not Alexandra.

- (6) And it cannot be a property of that person, since the resulting person will have the same essential and non-essential mental and physical properties whether she is Alexandra or not Alexandra.
- (7) So the distinguishing part must be a pure mental part, which we may call a “soul.”

But this argument fails. In order for it to work, we would have to assume that what in the passage quoted just above Swinburne calls the logical possibilities of the resulting person’s being Alexandra and of her not being Alexandra are also *real* (that is, what philosophers sometimes call “metaphysical”) possibilities—i.e., that the facts about the resulting person’s component physical parts and her mental and physical properties fail to settle the question of who she is, so that a person *really could* have had just these parts, and just these mental and physical properties, and either been Alexandra *or not*. If this were so, then what settled whether the resulting person was Alexandra would have to be something other than her component physical parts and her mental and physical properties. But anyone who holds a non-dualist theory of personal identity will deny just this: she will allow that though the sentences “The resulting person is Alexandra” and “The resulting person is not Alexandra” might both say things that are *logically* possible, since each of them “is not a contradiction and does not entail a contradiction” (14), nevertheless one of them says something that is a real (“metaphysical”) impossibility—since the resulting person must be either Alexandra or not, and whichever person she is, she is that person necessarily. And then the non-dualist philosopher can comfortably deny either or both of premises (5) and (6), saying instead that any person with *these* component physical parts, having this past history and arranged in this way, or any person with *these* essential and non-essential mental and physical properties, either will be Alexandra or she won’t be. Nothing extra is required to constitute the fact of the matter.

I conclude that it’s possible for a philosopher who holds a theory on which the identity of a person is fixed by the physical facts about him or her, so that a person *so* physically constituted, and with this past history, couldn’t be anyone other than who she is, to resist Swinburne’s argument for mind–body dualism. This is possible even if that philosopher accepts what I think we should not, viz. that the transplant scenarios I discussed in Section I are real possibilities, whose consequences would be what Swinburne supposes.

BIBLIOGRAPHY

- FOOT, Philippa. (2002). “The Problem of Abortion and the Doctrine of the Double Effect.” In *Virtues and Vices and Other Essays in Moral Philosophy*, 19–32. Oxford: OUP, 2002.
- MCDOWELL, John. “Reductionism and the First Person.” In *Reading Parfit*, edited by Jonathan Dancy, 230–50. Oxford: Blackwell, 1997.

SWINBURNE, Richard. *Are We Bodies or Souls?* Oxford: OUP, 2019.

THOMPSON, Evan, and Diego COSMELLI. "Brain in a Vat or Body in a World? Brainbound Versus Enactive Views of Experience." *Philosophical Topics* 39, no. 1 (2011): 163–80.

WILLIAMS, Bernard. "The Self and the Future." *Philosophical Review* 79, no. 2 (1970): 161–80.

NO WORK FOR A THEORY OF PERSONAL IDENTITY

S u m m a r y

A main element in Richard Swinburne's (2019) argument for substance dualism concerns the conditions of a person's continued existence over time. In this commentary I aim to question two things: first, whether the kind of imaginary cases that Swinburne relies on to make his case should be accorded the kind of weight he supposes; and second, whether philosophers should be concerned to give any substantial *theory*, of the sort that dualism and its competitors are apparently meant to provide, to explain the conditions of personal identity after all. My suggestion, instead, will be that the concept of a person's continued existence is better taken as philosophically unanalyzable.

Keywords: Swinburne; dualism; personal identity; imaginary cases.