

HOWARD ROBINSON

THE REVIVAL OF SUBSTANCE DUALISM

1. INTRODUCTION

Outside the philosophy of religion, where Richard Swinburne's influence has been enormous, his greatest achievement has been to put substance dualism back in play as an option in the philosophy of mind. I share his commitment to a belief in the mind as a simple mental substance, but my path to that conclusion is slightly different from his. I will argue in this essay that his revised version of Descartes' argument in chapter 5 of *Are We Bodies or Souls?*¹ does not quite get him to the conclusion that he requires, but that a modified version of his treatment of personal identity will do the trick. I will also look critically at his argument against epiphenomenalism, where, once again, I share his conclusion but have reservations about the argument.

2. WHY THE REVISED DESCARTES ARGUMENT DOES NOT WORK

Swinburne states Descartes' argument for the soul as follows:

first premise: I am a substance which is thinking.

second premise: it is conceivable that 'I am thinking and I have no body'.

third premise: it is not conceivable that 'I am thinking and I do not exist'.

lemma: I am a substance which, it is conceivable, can exist without a body.

conclusion: I am a soul, a substance, the essence of which is to think. (72–73)

HOWARD ROBINSON, University Professor of Philosophy at the Central European University; address for correspondence: Quellenstrasse 51, 1100 Wien, Austria; e-mail: robinson@ceu.edu; ORCID: <https://orcid.org/0000-0002-2972-5692>.

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

Swinburne thinks that this argument breaks down for a minor and for a major reason. The minor one is that the conclusion should be that he is a substance the essence of which is that he has a capacity to think, because he sometimes exists without thinking, for example, when unconscious. We will return to this later. The major reason is that the conclusion does not follow from the lemma:

But all that [the lemma] shows is that it is logically possible that he could “now” (that is, at the time when he is formulating his argument) be existing without a body and so with only a different part—a soul. It does not follow that he actually has a soul now. It shows only that if he has a soul, he exists; it does not show that he could not exist without a soul; that is, it shows that having a soul is sufficient for his existence, not that it is necessary for his existence. For maybe what is necessary and sufficient for Descartes to exist, what his whole essence consists in, is “to have either a body or a soul (or both).”

To assess this argument and the major objection, it is necessary to sort out what “conceivable” means. If it means “logically possible,” which Swinburne (73) says it does, then, if the lemma is correct, I think the conclusion follows. For if it is logically possible that I exist without a body, then I am not identical with or logically necessarily in any way related to my body, for if I weren’t necessarily related to it, it would not be logically possible that I’d be without it. So, if it is possible that I be without it cannot be that my essence is “to have either a body or a soul (or both),” for it cannot be to have a body at all.

To modify the jargon slightly, if it is logically possible that I could exist without a body, there is a possible world in which I exist without a body, so there can be no identity or other necessary or essential relation between me and a body.

So, if Descartes’ argument is to be unsound, it can only be because “conceivable” does not mean “logically possible,” but, say, epistemically possible; that is, as far as I can tell, I might be able to exist without my body.

In fact, what I think the argument shows is this: there logically could be somebody with a mental state just like mine, but lacked a body, but it is only epistemically possible that I might be such a person. We shall return to this thought, as, indeed, does Swinburne, I think.

Swinburne amends the argument by changing the second premise and adding a fourth:

first premise: I am a substance which is thinking.

amended second premise: it is conceivable that ‘while I am thinking my body is suddenly destroyed’.

third premise: it is not conceivable that ‘I am thinking and I do not exist’.

...

fourth premise: it is inconceivable that any substance can lose all its parts simultaneously and yet continue to exist.

amended conclusion: I am a soul, a substance, whose only essential property is the capacity for thought. (78–79)

The fourth premise is new, but uncontroversial. Swinburne's argument for the amended second premise is this:

If it is conceivable that while I am now thinking, I have no body, it is surely conceivable that while I am now thinking, I suddenly cease to have any control over my body or to be influenced by anything that happens in it; and in that case I would have ceased to 'have' a body; the body would be un-owned by me. (77)

First, the antecedent of the conditional the start of this passage is something that we have already shown to be false, so the argument does not start. Second, the sense of "ceased to have" is metaphorical in the case of loss of control or feeling. That I might not be dependent on my heart beating and my blood flowing to continue to exist is not shown by the loss of conscious control. Swinburne goes on to cite "out of the body," near death experiences to show what this might be like, but unless it is proved that these really are out of the body experiences, this does not seem to me to help. It might be replied that reliance on heartbeat and blood flow is just a causal reliance, but this begs the question. Again, as in the original version of the argument, it just shows that either it is epistemically possible, or that, in some other possible world, someone otherwise like me might be in this state.

At the end of chapter 4 Swinburne makes a concession that seems to me to be very similar to conceding my point against both versions, Descartes' and his:

It is normal for contemporary philosophers to claim that Descartes's argument fails for a reason quite different from those that lead to my amending the argument in the way that I discussed earlier in this chapter. Objectors normally acknowledge that the sentence 'I exist and have no body' does not by itself entail a contradiction and so is logically possible. But what that sentence tells us depends on who uses it. Descartes is using it to refer to the actual person referred to by 'I', himself, a person; and so what he is claiming is that 'I exist and have no body, given what "I" (as used by Descartes) refers to' is logically possible. If Descartes were a zombie, what he claims would be logically impossible. But the issue is whether it is possible, given Descartes is what he is. The way in which most contemporary philosophers express this point is to say that while the sentence may be *de dicto possible* ... the issue is whether it is *de re possible*. (84, emphasis in the original)

He concludes:

And of course, if this objection to Descartes's argument were correct, it would be an equally powerful objection against my own amended version of Descartes's argument. I will devote the next chapter to assessing this objection, and I shall conclude that each of us does know to what they are referring by 'I', and so we are in a position to assess what is logically possible for that substance to be or do. (85)

So, by the end of the chapter there is little disagreement between us, except that, on my way of expressing what is wrong with Descartes' argument, Swinburne's modification is not an improvement at all. To carry on the argument, we must move to chapter 5.

At the beginning of chapter 5, Swinburne, as a preliminary to proving that each subject has a grasp on his nature, distinguishes between what he calls *informative designators* and *uninformative designators*.

A word is an informative designator of some object iff the rules for its application to the object are such that speakers who know what the word means ... thereby know what is the essence of any object to which it applies (that is, a set of necessary and sufficient conditions for being that object). It is an uninformative designator if the criteria for its application to an object are such if it needs either further generally available knowledge including knowledge about the context of utterance..., or future scientific or other empirical investigation ... to determine which are the necessary and sufficient conditions for being that object. (94)

This distinction between informative and uninformative designators looks, at first sight, very similar to the distinction between transparent and topic neutral grasps on content. A topic neutral conception is one that picks something out by, for example, its role or function, but does not reveal its intrinsic nature. By contrast, a transparent conception involves a grasp on the intrinsic nature of the thing. The radical—and bizarre—nature of the physicalist theories of J. J. C. Smart and D. M. Armstrong consists in the claim that we have only a topic neutral grasp on the nature of the contents of our mental states. Our first-personal grasp on, for example, what it is like to see red, is only that it is some internal state, the nature of which we do not know, occurs when we face a red object, conditions are normal, etc. The contrasting "transparent" grasp on 'red' is that we ostensively define 'red' by what a certain quality is like in our experience. That is why how you understand 'red' could be what I take to be green.

Swinburne is surely right to reject the topic neutral account of our knowledge of the contents of our conscious states: we have a transparent or informative grasp on what it is like to see red, or what phenomenal red is like, which is the other side of the same coin. He wants to go further than this, however, and claim that we have a transparent or informative grasp on the subject or self that experiences or has these mental states.

Swinburne's argument for this conclusion is, it seems to me, inadequate. He argues that we cannot misidentify ourselves and that is enough for, or equivalent to, having an adequate or transparent grasp on what we are. It could be that he is misled by his own terminology here. Maybe it is enough for having an "informative designator" that you know which object you are referring to, but that does not seem to be the same as having a grasp on the intrinsic nature of the thing to which you are referring. Swinburne's concluding claim at the end of chapter 4 (cited above) was:

I shall conclude that each of us does know to what they are referring by 'I', and so we are in a position to assess what is logically possible for that substance to be or do. (85)

It is with the claim that it tells us what it is possible for that substance to *be* the kind of thing it is. Supposing physicalism were true, a conscious subject surely could know which subject he was. In fact, the problem here seems to be the same as it was at the end of the previous chapter: the subject knows who he is, but not fully what he is. I say "fully" because he does know that he is something with the capacity to be conscious, but he has known that all along. Taking the transparency/informative designator argument to have shown that the contents of our mental states are not physical states, because physicalism requires that our grasp on phenomenal contents be topic neutral/uninformative, does this give us grounds for concluding that the self is essentially and solely immaterial?

One reason for denying this would be a position adopted at one time by A. J. Ayer, namely that it is the fact all of a person's mental contents depend on the same brain that unifies them into the same mind. But it was pointed out, by John Foster, that dependence on the same brain did not constitute phenomenal unity, and Ayer accepted this criticism.

A slightly different line is to take the brain, or the human animal, to be the experiencing subject, and not merely the sustaining cause. But, *qua* physical, a brain or animal cannot possess the property of awareness of a non-physical state, as we have agreed, phenomenal contents are: the notion of a brain actually experiencing or apprehending something as the conscious subject, as opposed to sustaining such a thing does not make sense.

This shows that an immaterial subject is essential to a thinking being, but it does not ensure that it is the whole of the essence. The possibility of incompleteness is brought out by the fact that conscious subjects can undergo periods of unconsciousness. Descartes thought that consciousness continued dimly even in deepest sleep. I have defended the exotic view that, though the self acts within time, it actually exists outside it, so it need not ever be actually unconscious, only not be acting in the empirical world at certain times. Swinburne thinks, on the other hand, that it is enough that it is always has the *capacity* for consciousness. This answer may be

adequate, but it is plausible to argue that an unexercised capacity—one that is not actualized at a given time—must depend upon something actual which sustains its continuity. One could combine this idea with the “exotic” view by arguing that the identity of a soul *as the soul of a human person existing in the empirical order* depends on its association with a body. These seem to me to be open questions, even accepting the arguments above.

3. PERSONAL IDENTITY AND THE SIMPLE SELF

Swinburne is against all complex theories of personal identity because they allow indeterminate cases, which he thinks are unacceptable. For other objects—paradigmatically complex physical objects, such as Theseus’s ship—we are happy to say that it is a matter of convention or degree whether it be deemed the same object after radical repair, but Swinburne thinks this is unacceptable for personal identity. So, after partial brain transplants or other things that might seem to mix up identities, there must still be a definite answer as to whether someone survives. This, he argues, applies in the various kinds of case of brain-splitting and transplants that philosophers have imagined. There is no “partial survival” or “partial identity,” no “degrees of continuity.” The question “Will I survive?” must always have a definite “zero-sum” answer, even if one has no idea what it will be.

I do not think that this appeal to intuition will convince anyone tempted by complex physical or mental continuity theories. After the various operations have been carefully performed, splitting a brain and locating the two halves in different heads, it should be the case the resultant person or persons have a clear “unity of apperception” and so are like a normal, unified person in their consciousness at any given moment, but their sense of who they are, on reflection, with mixed up memories could be completely confused. It is not convincing that the original person should expect to be fully identified by either.

I have argued elsewhere² that it is counterfactuals of origin that show the need for the simplicity of the self, and not contestable cases involving identity through time. My latest statement on this is in my *From Knowledge Argument to Mental Substance*, and what I have to say is very similar to what I say there, in chapter 15.

² Howard ROBINSON, “Dualism,” in *The Blackwell Guide to the Philosophy of Mind*, ed. Stephen Stich and Ted A. Warfield (Malden, MA: Blackwell, 2003); ROBINSON, “Dualism,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, accessed September 25, 2020, <https://plato.stanford.edu/entries/dualism>; ROBINSON, *From Knowledge Argument to Mental Substance: Resurrecting the Mind* (Cambridge: CUP, 2016).

The standard “Theseus ship” cases concern repairs carried out in the course of the object’s existence, but we can also ask: Would it have been the same ship if it had been *originally made* with $n\%$ different planks? It is even clearer in this case than in change through time that there is no definite answer—it would have had a certain degree of overlap, and that is that.

Let us try to apply the same thought experiment to a human being. Suppose that a given human individual—call him Jones—had had origins different from those which he in fact had such that whether that difference affected who he was, was not intuitively obvious. We can approach this by imagining cases where it seems indefinite whether what was produced was the same body as Jones in fact possesses. What would count as such a case might be a matter of controversy, but there must be one. Perhaps it is unclear whether Jones’s mother would have given birth to the same human body if the same egg from which the Jones body came, had been fertilized by a different though genetically identical sperm from the same father. Some philosophers might regard it as obvious that sameness of sperm is essential to the identity of a human body. In that case, imagine that the sperm that fertilized the egg had differed in a few molecules from the way it actually was; would that be the same sperm? If one pursues the matter far enough there will be indeterminacy which will infect that of the resulting body. There must therefore be some difference such that neither natural language nor intuition tells us whether the difference alters the identity of the human body; a point, that is, where the question of whether we have the same body is not a matter of fact.

These are cases of substantial overlap of constitution in which that fact is the only bedrock fact in the case: there is no further fact about whether they are “really” the same object.

My claim is that no similar overlap of constitution can be applied to the counterfactual identity of minds. In Geoffrey Madell’s words: “But while my present body can thus have its partial counterpart in some possible world, my present consciousness cannot. Any present state of consciousness that I can imagine either is or is not mine. There is no question of degree here.”³

Why is this so? Imagine the case where we are not sure whether it would have been Jones’ body—and, hence, Jones—that would have been created by the slightly modified sperm and the same egg. Can we say, as we would for an object with no consciousness, that the story “something the same, something different” is the whole story: that overlap of constitution is all there is to it? For the Jones body as such, this approach would do as well as for any other physical object. But suppose Jones,

³ Geoffrey MADELL, *The Identity of the Self* (Edinburgh: Edinburgh University Press, 1981), 91.

in reflective mood, asks himself “if that had happened, would I have existed?” There are at least three answers he might give to himself. (i) “I either would or would not, but I cannot tell.” This is the option Swinburne defends, using the “change through time” cases. It is also the option I hope to show is correct, using a slightly different path. (ii) “In some ways, or to some degree, I would have, and in some ways, or to some degree, I would not. The creature who would have existed would have had a kind of overlap of psychic constitution and personal identity with me, rather in the way there would be overlap in the case of any other physical object.” (iii) “There is no fact of the matter whether I would or would not have existed: it is just a mis-posed question. There is not even a factual answer in terms of overlap of constitution.”

The second answer parallels the response we would give in the case of bodies. But as an account of the subjective situation, it makes no sense. Call the creature that would have emerged from the slightly modified sperm “Jones*.” Is the overlap suggestion that, just as, say 85% of Jones*’s body would have been identical with Jones’ original body, and about 85% of his psychic life would have been Jones’? That it would have been like Jones’—indeed that Jones* might have had a psychic life 100% like Jones’—makes perfect sense, but that he might have been to that degree, the same psyche—that Jones “85% existed”—makes no sense. Take the case in which Jones and Jones* have exactly similar lives throughout: which 85% of the 100% similar mental events do they share? Nor does it make sense to suggest that Jones might have participated in the whole of Jones*’s psychic life, but in a rather ghostly only-85%-there manner. Clearly, the notion of overlap of numerically identical psychic parts cannot be applied in the way that overlap of actual bodily part constitution quite unproblematically can. If one takes this view it has consequences for the actual case, not just the counterfactual. If you would have, in some way, partially existed if the sperm (or whatever) had been slightly different, all the people who would have existed under those various circumstances now partially exist in you. Something must have gone wrong with this thought!

The important thing about this argument in this context is how the identity across counterfactuals of origin case differs from that of identity through changes across time.

This concerns what one might call *empathetic distance*, which is essential to the problematic nature of identity through time but irrelevant in the counterfactual case.

Suppose that my parents had emigrated to China whilst my mother was pregnant with me, and that, shortly after my birth, both my parents had died. I was then taken in by Chinese foster parents, lived through the revolution and ended up being brought up in whatever way an alien would have been brought up in Mao’s China. None of this person’s post-uterine experiences would have been like mine. It seems, on the one hand, that this person would obviously have been me, and, on the other, that it is utterly unclear what kind of empathetic connection I can feel to this other

“me.” If I ask, like Jones: “Would this have been me?” I am divided between the conviction that, as the story is told, it obviously would, and a complete inability to feel myself into the position I would then have occupied. This kind of failure of empathy plays an important role in many stories that are meant to throw doubt on the absoluteness of personal identity. It is important to the attempt to throw doubt on whether I am the same person as I would become in fifty years time, or whether brain damage would render me “a different person” in more than a metaphorical sense. It is also obviously something that can be a matter of degree: some differences are more empathetically imaginable than others. In all these cases our intuitions are indecisive about the effect on identity. It is an important fact that problems of empathy play no role in the counterfactual argument. The person who would have existed if the sperm had been slightly different, could have had as exactly similar a psychic life to mine in as exactly similar environment as you care to imagine. This shows the difference between the cases I have discussed and the problematic cases that involve identity through time. In those cases the idea of “similar but not quite the same” gets empirical purchase. My future self feels, in his memory, much, but not all, of what I now feel. In these cases, overlap of conscious constitution is clearly intelligible. But in the counterfactual cases, imaginative or empathetic distance plays no essential role, and the accompanying relativity of identification gets no grip.

Option (iii) is more subtle, or, perhaps, more mysterious. Is that conception of the self which makes us feel so sure that someone psychically just like me but with a somewhat different origin either is me or is not something that needs “deconstructing,” after the fashion of Derrida or Nietzsche or Hume?

I do not think that the idea “just like me but the idea of whether it would be me or not has no content” can be made acceptable. Whereas in the case of physical objects we can see, after a little thought, that though the qualitatively similar gives us all we thought we needed by talking about particulars, it will not do this in the case of minds.

Consider the following example. Suppose you discover that, in the very early stages in the womb, you were one of twins, but that the other did not develop, and that it could have easily happened the other way round; the other would have survived and you died in the first few days. The similarity between you as survivor and your twin, had he survived in your stead, both in genetic endowment and environmental circumstances and subsequent experience, could have been almost complete. Nevertheless, there is no sense that, on reflection, it makes no serious factual difference, concerning your own fate, which of the two survived. Just as it is true that, if your parents had never met, then you would not have existed, equally, if the other bundle of cells had developed instead of yours, you would not have existed. This is, in no sense, a matter of decision, convention or degree.

It is possible to answer this argument by saying that the twins' case is simply a matter of two different bodies, and I do not deny that this difference can be taken as foundational. The case is an analogy to show that there really is something at stake and is a response to the "no fact of the matter" strategy. The "different bodies" response is explaining the supervenience base for saying it would not be me. And saying it would not be me is not just to say it would be a different body—it is something real that goes along with its being a different body. (Perhaps the situation is not even that simple: both came from the same zygote and so sameness of body might be an open issue, given that there never developed an alternative personal claimant.)

It would mean that the idea that it would be you, if counterfactual *p* is never any different from the proposition that there would have been someone like you if *p*, whatever *p* may be. It could be that someone cough on the other side of the world at the moment of your conception. We certainly cannot let go the idea that there is genuine and primitive sameness under these circumstances.

It would seem that the only possible answer to the question which I supposed Jones to have asked himself above, namely "if that had happened, would I have existed?" is (i), "I either would or would not, but I cannot tell." If there is a real fact, independent of our convention or decision, in this case, then it shows that counterfactual identity facts are real facts in the case of minds, in a way that they may not be for physical objects.⁴

4. CONCLUSION

I think that this book opens up many issues concerning substance dualism in very valuable ways. In particular, I believe that Richard Swinburne is right in defending the claim that the self must be simple and, hence, immaterial (given that the idea that it is some strange material particle is bizarre, if intelligible) is correct. But I think that the "counterfactuals of origin" argument is more conclusive on this point than either a version of Descartes' argument, or questions of identity through time.

There is, however, something that those of us who believe in mental substance need to say more about, namely the contribution of the immaterial self to the personhood and psychology of human beings. It can seem from the arguments presented

⁴ One way, of course, of avoiding the conclusions of this section would be to adopt a Lewisian theory of counterfactuals, according to which there is no true counterfactual identity, only counterparts. This would be a heavy price to pay. I have a little to say about this and of non-realist accounts of counterfactuals in section 15.5 of my *From Knowledge Argument to Mental Substance*.

that all that it contributes is a bare “thisness” or identity, and nothing to any richer sense of “who we are” or how we act as free, moral beings and intelligences. It can sometimes seem as if we accept that all this can be put upon the brain and other aspects of physical endowment. We need more on what one might call the philosophical psychology of substance dualism, and of how embodiment—the real issue of “the mind–body problem”—might work.⁵ This connects with my remarks at the end of section 2, concerning the role of the body in creating our identity as human within the empirical order, and, maybe, is also a modest gesture towards a more Thomist picture of the relation of body and soul.

BIBLIOGRAPHY

- MADELL, Geoffrey. *The Identity of the Self*. Edinburgh: Edinburgh University Press, 1981.
- ROBINSON, Howard. “A Dualist Account of Embodiment.” In *The Case for Dualism*, edited by John R. Smythies and John Beloff, 43–57. Charlottesville: University of Virginia Press, 1989.
- ROBINSON, Howard. “Dualism.” In *The Blackwell Guide to the Philosophy of Mind*, edited by Stephen Stich and Ted A. Warfield. Malden, MA: Blackwell, 2003.
- ROBINSON, Howard. “Dualism.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed September 25, 2020. <https://plato.stanford.edu/entries/dualism>.
- ROBINSON, Howard. “Substance Dualism and Its Rationale.” In *Free Will and Modern Science*, edited by Richard Swinburne, 158–77. Oxford: OUP (for the British Academy), 2011.
- ROBINSON, Howard. *From Knowledge Argument to Mental Substance: Resurrecting the Mind*. Cambridge: CUP, 2016.
- SWINBURNE, Richard. *Are We Bodies or Souls?*. Oxford: OUP, 2019.

THE REVIVAL OF SUBSTANCE DUALISM

Summary

I argue in this essay that Richard Swinburne’s revised version of Descartes’ argument in chapter 5 of his *Are We Bodies or Souls?* does not quite get him to the conclusion that he requires, but that a modified version of his treatment of personal identity will do the trick. I will also look critically at his argument against epiphenomenalism, where, once again, I share his conclusion but have reservations about the argument.

Keywords: Swinburne; dualism; personal identity; epiphenomenalism.

⁵ For a modest start, see Howard ROBINSON, “A Dualist Account of Embodiment,” in *The Case for Dualism*, ed. John R. Smythies and John Beloff (Charlottesville: Virginia University Press, 1989), 43–57.