

the transplant argument

—

Dustin Crummett, LMU Munich

the transplant argument

This is an argument (inspired by Swinburne, modified by me) *for* the claim that I am a simple, indivisible, immaterial soul and *against* the claim that I am a material object. I have simplified the argument in some ways for present purposes.

If (as I think) the argument establishes that the connection between my mind and body is contingent, then it works just as well against the claim that the mental *necessarily supervenes* upon the physical (i.e., that there can be no mental differences without physical differences). I think Ben will endorse the supervenience claim.

subjecthood

Experiences don't just float around; they are had *by* experiencers. I am such an experiencer--I am a *subject of experience*. When I was a little fellow, I was very different than I am now. But the subject of those experiences is the same as the subject of the experiences I am having now.

Suppose someone on Twin Earth has, for their entire life, had the exact same experiences as me. There is an important difference between these two streams of consciousness: one was had by that person, and one was had by me. (If our experiences differ in the future, which stream is which might be really important to me.)

materialism

If I am material, presumably I am my brain, or my body, or something fancier (an event, an object constituted by but not identical to my brain, etc.) For simplicity's sake, suppose the materialist thinks I am my brain. But this shouldn't matter.

I want to argue that no view on which I am a material object can adequately account for the way in which subjects of experience are simple and indivisible. So no subjects of experience are material, and, *a fortiori*, neither are you or I.

hemispherectomy

Sometimes people have a hemisphere of their brain removed for medical reasons. It turns out that enough functions (memory, personality, etc.) are duplicated between hemispheres, and that there is enough neuroplasticity, that you can have this surgery and (potentially) not really change very much.

It seems that I can survive a successful hemispherectomy, in the sense that the subject of experience *before* the surgery is the same as the subject of experience *after* the surgery. (Or at least it's metaphysically possible that I do. This is enough for the argument: just assume that, when I speak as if I *actually* survive, I'm speaking about the worlds where I do survive.)

single transplant

Consider a case called *single transplant*. Suppose I have a hemispherectomy. The removed hemisphere is destroyed. Some time later, my body is failing. Futuristic scientists successfully transplant my remaining hemisphere into a new body. My old body dies.

I think I survive this, too. I “go where my brain goes,” so to speak: I am the person in the new body. After all, (the relevant part of) my brain is the causal basis of my thoughts, memories, experiences, etc. (Or at least, it’s metaphysically possible that I survive.)

single transplant, no wait

Now suppose we do the hemispherectomy, removing one hemisphere and destroying it, while *at the same time* removing the other hemisphere and transplanting it. This is just *single transplant*, except we don't wait before doing the transplant.

If I can survive in *single transplant*, I think I can survive here. Why would waiting to do the transplant make any difference? (At least, it's metaphysically possible, etc.)

double transplant

But consider *double transplant*. *Both* hemispheres of my brain are simultaneously removed *and transplanted*.

The left hemisphere is put into a new body which is identical to my old body. Call the resulting person *Mr. Left*.

The same thing happens with my right hemisphere. Call this person *Mr. Right*.

who am I?

Mr. Left and Mr. Right are both physically and psychologically continuous with me before the surgery. But they cannot both be me. Mr. Left is in this room and about to do one thing, and Mr. Right is in that other room and about to do a different thing.

So either (i) I am Mr. Left, and Mr. Right is a new person, or (ii) I am Mr. Right, and Mr. Left is a new person, or (iii) I died, and Mr. Left and Mr. Right are both new people.

my view

My view is: I am an immaterial soul causally related to a material body. The laws governing this relationship are contingent. Because we do not know the exact form of these laws, we don't know which of (i), (ii), or (iii) obtains. The answer is genuinely contingent.

Since these laws are highly complex and arbitrary, it's implausible to posit them as a brute fact. They are too intrinsically improbable. We should instead seek to explain them in terms of something deeper. I explain them in terms of divine action. (We can say the same about the laws linking conscious and physical states in general.)

the argument

Why accept this? Whatever we say happens to my brain (or any other material object I might be) in *double transplant* is not something we can plausibly say happens to me. So I am not my brain (or any other material object), by Leibniz's law.

Roughly, there are four possibilities for what happens to my brain.

deep, determinate fact: mereological nihilism

Maybe there is a deep, determinate fact about what happened to my brain: it never existed, because mereological nihilism is true.

(I actually think this.)

But then I'm obviously not my brain, since I exist and my brain doesn't.

partial identity

We might say my brain is *partly* identical to Mr. Left's, and *partly* identical to Mr. Right's brain. This seems plausible enough.

But if Mr. Left and Mr. Right are also their brains, then I am partly identical to Mr. Left and partly identical to Mr. Right. And that can't be true, thinking of all of us as subjects of experience. If Mr. Left is tortured and Mr. Right eats ice cream, there is no one who is 50% the experiencer of the torture and 50% the experiencer of the ice cream taste. (What would that even mean?) Sameness of subject is “all or nothing” in a way incompatible with this proposal.

no deep fact

We might say there is no deep fact of the matter about whether my brain is Mr. Left's or Mr. Right's or neither's. Maybe it's indeterminate. Or maybe it's just a matter of convention or stipulation.

But that can't be true of me. There is some fact of the matter about whether I am having conscious experiences or not. And if I am, there is some fact of the matter about whether they are Mr. Left's or Mr. Right's. Further, this is not a matter of convention or stipulation: if, after the surgery, Mr. Left will be tortured while Mr. Right gets ice cream, it does me no good to just *decide* that I'll be Mr. Right.

deep, determinate fact: composition

Finally, maybe there is a deep, determinate, non-partial-identity-involving fact about what happened to my brain, and it did exist. Either: (a) my brain stopped existing (two new brains are created), or (b) my brain is Mr. Left's (Mr. Right's brain is new), or (c) my brain is Mr. Right's (Mr. Left's brain is new).

If this is right, I have two arguments against identifying this with my brain.

modal

It seems to me that it is *genuinely metaphysically possible* that I could be Mr. Left, or Mr. Right, or die, holding all the physical facts fixed. However, the identity of a material object can't change unless something else in the physical world changes (the identities of material objects supervene upon fundamental physical facts). E.g., there can't be two worlds where the *only difference* is whether my table is the same table it was yesterday: something else has to be different (e.g., the table getting or not getting a new part) to explain that difference.

So if all these scenarios are metaphysically possible, I must be non-physical, or have some non-physical part which does something different in different scenarios.

modal

Why accept that it's metaphysically possible that I be any of the three?
Conceivability.

Possible objection: conceivability does not entail possibility.

Response:

- 1) Conceivability is at least *evidence* for possibility.
- 2) Where I posit a contingent law, the materialist posits a (non-self-evident) necessary connection. But the contingent law can be explained by divine action. Absent some explanation of the alleged necessity, I eliminate an unexplained and arbitrary posit. (I claim this point holds for mental/physical connections more broadly).

intrinsicity

Finally, suppose we grant that I survive (or at least that it's metaphysically possible for me to survive) in *single transplant, no wait*. This means that, if I am my brain and the view in question is true, whether an object is identical to me is an *extrinsic* matter: you can change whether an object x is *me* merely by changing other things, without changing x itself in any way. But surely this is not correct.

intrinsicity

Here is why the view entails extrinsicity:

Suppose we say I am neither Mr. Left nor Mr. Right. But if we had destroyed my right hemisphere while implanting my left hemisphere in Mr. Left's body, I would have been Mr. Left's brain (by *single transplant, no wait*). So whether I am Mr. Left's brain is determined by what happens to the other hemisphere in the other room, without affecting Mr. Left's brain in any way.

intrinsicity

Suppose we say I am Mr. Left. But if we had destroyed my left hemisphere while implanting my right hemisphere in Mr. Right's body, I would have been Mr. Right's brain (by *single transplant, no wait*). So whether I am Mr. Right's brain is determined by what happens to the other hemisphere in the other room, without affecting Mr. Right's brain in any way.

Same goes, necessary changes being made, if we say I am Mr. Right.

intrinsicity

The soul view avoids this: I am my soul no matter what. What happens to the brain in the other room might affect *which body I am causally related to*, if any. But of course what I am causally related to can be affected by things outside myself, so it isn't a problem.

intrinsicity

Intrinsicity is also a worry for partial identity and no deep fact: they make it possible to change whether an object x is *wholly, determinately* identical to me by changing factors outside of it, without changing x itself in any way. (Proof left as an exercise for the reader.)