

Explaining why this body gives rise to me *qua* subject instead of someone else: an argument for classical substance dualism

KENNETH EINAR HIMMA

*Department of Philosophy, Seattle Pacific University, 3307 Third Avenue West,
Seattle, WA 98119
e-mail: himma@spu.edu*

Abstract: Since something cannot be conscious without being a conscious subject, a complete physicalist explanation of consciousness must resolve an issue first raised by Thomas Nagel, namely to explain why a particular mass of atoms that comprises my body gives rise to me as conscious subject, rather than someone else. In this essay, I describe a thought-experiment that suggests that physicalism lacks the resources to address Nagel's question and seems to pose a counter-example to any form of non-reductive physicalism relying on the mind–body supervenience thesis, which would include William Hasker's emergent dualism. Since the particular thought-experiment does not pose any problems for classical substance dualism (CSD) and since the problem, as I call it, of explaining subjectivity is the central problem of mind, I conclude that CSD is better supported than any form of non-reductive physicalism.

Introduction

Every conscious mental experience involves two constituents that seem to require explanation: (1) awareness of particular content in the form of a thought, belief, feeling, or perception (2) instantiated by a particular conscious mental subject. Since something cannot be conscious without being a conscious subject, a complete physicalist explanation of consciousness must explain a number of facts about conscious subjects entirely in terms of the causal properties of physical entities, states, and processes. In particular, it must address an issue first raised by Thomas Nagel,¹ namely to explain why a particular mass of atoms that comprise my body gives rise to me *qua* conscious subject, rather than someone else.

In this essay, I describe a thought-experiment that seems to show that physicalism lacks the resources to provide such an explanation while classical

substance dualism (CSD) has such resources. The thought-experiment thus calls into question the idea that mental states supervene on neurophysiological states and hence challenge both non-reductive physicalism and emergent dualism insofar as they assume this idea. I will consider a number of counter-arguments to this thought-experiment and argue that they fail.²

The notion of a conscious subject

Being a conscious subject is itself an element of conscious experience. We are conscious subjects with conscious mental states and the two are related in an intimate way. It seems clear, as a conceptual or metaphysical matter, that only conscious mental subjects can have *conscious* mental states of any kind. The idea that there are, so to speak, free-floating *conscious* mental states not experienced by some mental subject seems incoherent: it does not seem possible for a conscious mental state to occur that is not had by a mental subject. Conscious mental states or events can happen only to conscious subjects because they are the kind of thing that must have a bearer and subjects are, by definition, the bearer of such states.

Moreover, being a mental subject is the most basic element of consciousness. Something that functions as or is a subject of conscious states is itself a particular locus or centre of awareness and is itself conscious in a limited way. As J. P. Moreland and William Lane Craig describe this element of consciousness:

When we pay attention to our own consciousness, we can become aware of a very basic fact presented to us. We are aware of our own self as being distinct from our bodies and from any particular mental experience we have. We simply have a basic, direct awareness of the fact that we are not identical to our bodies or our mental events; rather we are the selves that *have* a body and a conscious mental life.³

Although Moreland and Craig intend this as an argument for CSD, the thesis that human persons are composite of material bodies and unextended, immaterial, mental souls that constitute a person as a conscious subject capable of experiencing mental content, I intend the passage above only as support for the intuition that there are conscious subjects that are metaphysically distinct from the content they instantiate (e.g. beliefs, desires, and perceptions).⁴

Indeed, as Moreland and Craig suggest, what I take *to be me* at the most intimate level is a particular locus of consciousness, which is the subject of all my conscious experience, including the indirect ones of my own subjectivity. Independent of all perceptual and semantic mental content, I am this locus of consciousness that is subject to all other mental content I experience – though what the content of this more basic awareness is like is difficult to say. As long as I am alive, and the higher parts of my brain are functioning properly, I am no less conscious if I am denied perceptual content somehow and am somehow able to quiet the inner voice that chatters incessantly throughout the day. Perhaps the

awareness in such cases is of nothing more than consciousness, but consciousness, the state of being aware, is surely something that has a feel – and my being the one experiencing that feel will presumably be part of the experience of pure consciousness.

Conscious subjectivity is perhaps the most basic element of any conscious experience. Subjectivity is a particular locus or centre of awareness and hence involves awareness. Other mental states involve awareness of particular kinds of content that are not involved in simply being aware or conscious *simpliciter* – whatever this turns out to be. We have a self-awareness (i.e. an awareness of ourselves as conscious subjects) that is difficult to characterize; it is not clear whether this involves some sort of direct awareness of subjectivity itself or whether it is some other mental state that is indirectly aware of subjectivity. But, either way, the state of being a conscious subject necessarily involves the state of being conscious – and this is a metaphysical (and probably logical) presupposition for having mental states that involve awareness of other forms of content, like perceptions, etc.

But it is important to note the crucial link between the particular locus of consciousness and what we take to be our selves in the most basic sense. If my body dies while the locus of brute consciousness that was subject to the mental experiences that came with that particular body continues to survive, then I have survived my death in the sense that is most important to me. I survive as a conscious, though not necessarily sentient, being. The content of my mental life might be quite simplistic and uninteresting, and my personality traits and memories might die with the body, but I will have, in some sense that matters deeply to me, survived my death.

The problem for physicalism

It has been difficult to express the foundation of physicalism in terms of non-physical/physical supervenience theses,⁵ but the consensus is that some form of supervenience thesis is partly (though not necessarily wholly) constitutive of that foundation. For our purposes, I think the following formulation is fairly characterized as a necessary, though not sufficient, member of the set of propositions common to all forms of physicalism:

The mind–body supervenience thesis: There cannot be a mental difference in this world without an underlying neurophysiological difference.⁶

Thus construed, the mind–body supervenience thesis is interpreted as a theory of mind, and not as an ontological claim equivalent to that made by materialism, which would preclude there being a non-physical being.⁷ The claim roughly amounts to the claim that the content of any mental state is fully determined by

neurophysiology. This should not be taken to deny that environmental factors play a role in determining the content of mental states and processes, but they do so only insofar as they affect the state which a subject's brain and nervous system is in.

Given that being a subject is a constituent of consciousness, the physicalist must explain it in terms of neurophysiological states and processes. One problem involves identifying the neurophysiological states that correlate with being a subject (or particular centre of brute conscious awareness); this involves the sort of empirical analysis that attempts to map particular items of conscious experience onto neurophysiological states. Another is a more intuitively disturbing problem involving mental subjects. As Thomas Nagel describes this problem:

It isn't easy to absorb the fact that I am contained in the world at all. It seems outlandish that the centerless universe, in all its spatiotemporal immensity, should have produced me, of all people – and produced me by producing TN [i.e. Thomas Nagel]. There was no such thing as me for ages, but with the formation of a particular physical organism at a particular place and time, suddenly there *is* me, for as long as the organism survives. In the objective flow of the cosmos this subjectively (to me!) stupendous event produces hardly a ripple. How can the existence of one member of the species have this remarkable consequence? (*VFN*, 55)⁸

The problem for the physicalist is to explain how it is that the particular body that was born at a particular set of points in space-time (i.e. the first one born to my mother) creates *me* as a particular subject – and not someone else. In other words, the physicalist must explain how, so to speak, the set of mereological simples arranged in the form of my body – or, more specifically, my brain – brings *me* into existence *qua* conscious subject, rather than someone else.⁹

The possibility of a perfect twin

Here is a thought-experiment that calls attention to a dilemma Nagel's problem poses for physicalism by suggesting a possible mental difference of the sort physicalism is obliged to explain without a neurophysiological difference to explain it. Suppose the earth has a perfect twin *in this universe* distinguishable from the earth only in terms of spatial and physical properties not known to anyone on either planet – though they would be known to someone with a God's-eye view of the universe.¹⁰ In this world, then, you have a twin who, though composed of different material than you, is in every other respect that can be discerned by us, mentally and physiologically, indistinguishable from you.

Apart from being composed of different materials (and the differences entailed by this), you and your twin are indistinguishable in terms of physical characteristics. You are genetically indistinguishable. You are physiologically isomorphic in the following respect: at every moment in time, the two of you have materially distinct but otherwise indistinguishable atoms and molecules arranged according

to the same blueprint, and those materials are always in exactly similar physical states. Indeed, you and your twin always agree on all subatomic events – including those that might lack an explanation in terms of paradigmatically causal relations. Your bodies are thus always in perfectly *isomorphic* states (though they are obviously not physically *identical*) at every level of description – indeed, all the way down to the sub-atomic level.¹¹

Likewise, you and your twin's mental states and characteristics track each other at every moment in your lives. You and your twin are exposed to exactly similar – though obviously not the *same* – sensory input at all times, and your brains respond to this input in qualitatively indistinguishable ways.¹² The brain-states associated with long- and short-term memory are indistinguishable in terms of their neurophysiological features; hence you and your twin have exactly similar long- and short-term memories at every moment in time. The brain-states associated with personality and emotional characteristics are likewise indistinguishable; hence you and your twin have exactly similar personality and emotional characteristics at every moment of your lives. Indeed, if we assume that we have libertarian free will, you and your twin will always instantiate exactly similar volitions at the same moments in time – even on the assumption that volitions are utterly uncaused.¹³

But while you and your twin, then, are mentally and physiologically indistinguishable at every level of description (again, apart from the spatially distinct materials that compose your bodies), there is one difference between you and your twin that seems to be very significant from an intuitive point of view: one of these subjects is *you* and the other is not. You are the subject of a stream of experience that arises from one of these two perfectly similar bodies with perfectly similar histories and not the other. That is, the subject of your mental life that you identify as *you* is paired with the content and *qualia* that arises from one of these bodies and not the other body that is qualitatively indistinguishable with respect to any neurophysiological state, function, or operation that might be relevant in explaining conscious experience.

At first glance, this poses a problem for any non-reductivist theory grounded in the mind–body supervenience thesis, including emergent dualism (since it holds that souls ‘emerge’ from neurophysiological processes).¹⁴ The thought-experiment seems to show a mental difference (one of the subjects is *you* and the other is someone else) that a theory of mind is obliged to explain if Nagel is correct in seeing a genuine problem here without a corresponding difference in neurophysiological states. If this is correct, then it is false that there can be no mental difference without a neurophysiological difference, which is what the mind–body supervenience thesis states. (It should be clear that spatial differences are relevant in explaining differences in mental states only insofar as they produce neurophysiological states that are different in terms of their neurophysiological operations and component-states.)¹⁵

Indeed, if the difference between subjects is one a theory of mind is obliged to explain, such considerations seem to show that there is no non-reductive physicalist explanation as to why *you* happen to experience the stream of consciousness that is brought about by physical stimulation of your body rather than the stream of consciousness that is brought about by *any other living body* sophisticated enough to give rise to conscious experience. From the standpoint of non-reductive physicalism, it is utterly arbitrary that you are the subject of a human being's conscious mental states; as far as the physical laws of this universe are concerned, the subject of your experience might very well have been paired with the conscious mental life, such as it is, furnished by a cat's body.

No similar problem is posed for CSD. The difference between me and my twin is that we are different conscious subjects and that difference is wholly explained by the thesis that we have different souls. Although empirical studies pairing mental and neurophysiological states have whittled down the number of mental states, functions, and processes that can be attributed as directly performed by the soul independently of neurophysiology, one job my soul does (and possibly the only one) is that it is the unique locus of consciousness that is me *qua* conscious subject and the bearer of all my mental states, regardless of ultimate source. Unlike emergent dualism, CSD does not claim that souls arise or emerge from physical processes: the existence of souls, according to CSD, is both metaphysically and nomologically independent of the existence of any particular physical organisms.

Indeed, this thesis of CSD coheres nicely with the common intuition that people have that they could have been, so to speak, in someone else's body. For example, I have the intuition that I could have 'inhabited' Ronald Reagan's body. This thought simply amounts to the idea that the subject of my experience might have been paired with the content-parts to which Ronald Reagan's body gave rise – rather than with the subject that Ronald Reagan identified as being himself. Of course, had the subject of my experiences been paired with the content-part of his experiences, I (*qua* subject) would have had Reagan's memories and personality (and hence would have been a different 'person' in the sense with which the theory of personal identity is concerned). Moreover, having died in 2004, I would no longer be the subject of any perceptual experiences – at least not any associated with the operation of sense organs in Reagan's body. But all those things, on this not uncommon intuition, would have been instantiated by a different conscious subject – namely, the one that is defined by the subject of my mental experience.

In any event, while there are a host of problems with CSD (e.g. how to differentiate souls), the twin-earth thought-experiment above poses no problem for the somewhat unfairly discredited theory. Indeed, the thought-experiment above seems to entail the truth of either some intuitively implausible reductivist physicalist theory or CSD.

Is the difference between me *qua* subject and my twin *qua* subject simply historical?

To defend the supervenience thesis, the physicalist must make the following claim: there is no difference between me and my twin conceived as just subjects – or, for that matter, me and any other person considered simply as a subject of experience (henceforth I will describe this idea as ‘*qua* subject’ or ‘*qua* centre of consciousness’) – that a theory of mind is obliged to explain. That is, the physicalist must claim that my twin and I, considered as just particular centres of consciousness, are exactly similar in all the ways relevant to a theoretical explanation of mind.

To see why someone might be tempted to think there is no theoretically significant difference between me and my twin *qua* subjects, it is helpful to start with something as simple as a table. Surely, a global physicalist theory¹⁶ of tables is true; every property, state, or operation (assuming there are any) of a table can be fully explained by a story that makes reference to only the physical constituents of tables and the nomological laws governing those tables. But the table at which I am currently sitting has a twin in the thought-experiment and is hence indistinguishable in every relevant respect to this one, but we do not have the intuition that there is something left to be explained – i.e. why the table that I am sitting at is not the table that my twin is sitting at.

Of course, the two tables have, strictly speaking, different histories in the sense that they originate from different trees; however, the trees are the same with respect to all their physical characteristics. The different histories can be explained easily in physicalist terms: the two physical tables, though indistinguishable in terms of what can be observed at any given moment, are causally linked to different physical objects – all the way back to the seeds and their different histories. Beyond this, there is nothing left to be explained by a physicalist theory. All properties of the table can be explained in purely physicalistic terms – as long as the relevant supervenience thesis can be expressed broadly enough to include the causal connections that constitute historical connections, which I assume can be done.¹⁷

One might think that the question Nagel asks is analogous to the question of why the table I am sitting at is not the table that my twin is sitting at. The question of why I am the subject associated with this mass of mereological simples and not some other mass, like those constituting my twin or, for that matter, those constituting you, poses no dilemma for physicalism because there are no mental differences that need to be explained in terms of some neurophysiological difference. Whatever mental difference there might be can be explained in terms of different histories: the subject that is you are associated with one body because of its history, while the subject that is your twin is associated with the other because of its history.

Notice that if the only difference between two streams of brute awareness is history, then the example of me and my twin would be very much like the example of the two tables. In fact, it would have an easier physicalist explanation than any other aspect of consciousness: any neurophysiological state sufficient to explain *any* piece of conscious mental content would be enough to explain this brute state of awareness that is qualitatively indistinguishable from one subject to another because every brain state that produces an awareness of some content must also produce awareness *simpliciter* – since one cannot be aware of something without being aware.

Perhaps one might reasonably think that there is no significant non-historical difference in mental states involving no more than a brute awareness without any extraneous (or extrinsic) mental content. The state of my twin's being brutally aware might be indistinguishable from the state of my being brutally aware in terms of content, *qualia*, or whatever features of brute awareness need to be explained by a theory of mind. I have, of course, no way of knowing whether this is right or not, but it doesn't seem wildly implausible.

Of course, if simply being aware and thus being capable of being aware of extrinsic content were all there is to being a particular subject, then Nagel's problem would, again, be trivially easy to solve – and it would hardly be a remarkable fact that for all the years TN's body did not exist, TN did not exist as a subject of consciousness. That this sort of reading of Nagel turns what seems plausibly to state a genuine difficulty for any philosophy of mind becomes a trivial problem on this view of the nature of being a subject seems to suggest that something has gone wrong, since virtually no-one prior to some significant theorizing would reject Nagel's description as stating a genuine problem.

But there is more to being a subject than just being aware. A subject is fairly characterized as a conscious *someone* (in the weakest sense) who is conscious in virtue of being brutally aware and hence capable of being aware of extrinsic content. Brute awareness is not just free-floating mental content – say, a free-floating state of awareness *simpliciter*; it is content that must be instantiated by someone, whatever a someone turns out to be – whether entity, substance, or property. And it is this fact that being a subject entails being some particular entity that is brutally aware that, from the standpoint of ordinary intuition, seems to make a difference that requires theoretical explanation that goes beyond citing different histories.

In particular, I do not have any reason to care about whether I instantiate one or the other of these brute awarenesses or sensations of being aware of being aware – whether peripherally or directly (and both are possible depending on how one shifts one's inner focus) – but I do have a special reason to care about the *someone that I am* who instantiates that content. The continuing existence of me *qua* subject matters much more to me from a practical standpoint than the continuing existence of my twin *qua* subject. In contrast, I have no reason to

care about whether I *qua* subject instantiate one or the other of two states of brute awareness (assuming they are distinguishable only in terms of historical considerations). One qualitatively indistinguishable stream of brute or self-awareness is as good, by my lights, as any other, provided that I exist *qua* conscious subject to instantiate it.

And, as far as I can tell, this seems to be true of any mental *quale* or content that is indistinguishable from any other on any grounds other than historical. I am utterly indifferent, for example, as to which of two qualitatively indistinguishable tastes of my favourite food is mine, or which of two qualitatively indistinguishable orgasms is mine, or which of two beliefs of exactly the same content is mine. Since my twin and I agree on all this *qualia* and content, I couldn't care less which particular *quale*-state or content-state is mine.

This appears to be true as a general matter with respect to rational preferences of any kind. It is irrational to prefer some item A to item B *unless there is some qualitative difference between A and B* and one prefers the relevant quality of A to the relevant quality of B. If A and B are utterly indistinguishable in terms of their qualities, one has no reason to prefer one to the other.¹⁸ From the standpoint of practical rationality, there are standards that govern which kinds of differences make a difference with respect to one's preferences.

And herein lies the difference between Nagel's problem and the problem posed by the tables. While it might be true that there is no relevant difference between the tables here and on twin earth that would explain a preference for one over the other, it simply does not seem true that there is no such difference between the subject of my twin's brute awareness and the subject of my own. Again, the streams of consciousness, considered by themselves, might very well be indistinguishable in every way that matters with respect to our preferences – along with the brute awareness of every sentient being; perhaps consciousness is that kind of thing.

But subjects experience streams of consciousness, and every two subjects are different in a way that makes a supreme difference when it comes to certain kinds of intuition. If the histories of my planet and its counterpart diverge at some point and my mother never gives birth to a son but my twin's mother does, I never come into existence while my twin does. All these humans in the universe, and I am nowhere to be found; but should my mother give birth to the son she gave birth to in September 1957, I come into existence.

This is the remarkable fact about my existing as a subject of experience that is plausibly claimed by Nagel to need explanation. And this fact is 'remarkable' because of its great importance to us from the standpoint of practical rationality. No psychologically, physically healthy person wants to die – even if death doesn't involve the threat of eternal punishment and simply involves the cessation of conscious experience; in fact, psychologically, physically healthy persons usually want very much not to die. This is, of course, partly explained by the fact that

conscious experience is frequently fun: food, sex, friendship, art, comedy, etc. are all goods that make conscious experience worth having, and hence life worth living. But this seems also partly explained by the fact that we value *our* continuing sentient existence as an end-in-itself – and not merely as a means to experiencing the mental states that enable us to experience those goods of life. As long as my life is not intolerably bad (e.g. I am not being tortured or have a terminal disease involving terrible pain that cannot be alleviated), I want the subject of my experience to continue to exist – indeed, more than I want just about anything else. Indeed, this desire is expressed in (and, as many would have it, explained by) the urgency of this desire to persist indefinitely as a subject of conscious experience.

Admittedly, there is much mysterious about what it is to be one particular subject rather than another. It is not clear, and seems somewhat odd to think, that the experience of my twin as just a conscious someone has a different *qualia* from mine – at least not in the sense that the experience of red has a different *qualia* than the experience of blue. If it did, then the problem the thought-experiment poses for physicalism would be obvious: we have a difference in *qualia* without a neurophysiological difference – and this is inconsistent with the mind–body supervenience theory.

Whatever the proper characterization of this difference is, it is hard to make sense of our strong preferences regarding our persistence as subjects just in terms of historical differences. My survival in the sense that matters most depends on the survival of that which functions as the *subject* of my conscious life, and I care about that more than I care about the survival of my body. *If* I could survive as a subject without my body's surviving, the survival of that subject would matter more than the survival of this body – though I certainly care about the survival of this body, in part, because of what it knows, which is stored in, and made possible by, having a brain with a particular causal history.

Differences in history are sometimes relevant, as we have seen, in explaining preferences for objects that differ in no other way, as was true of the tables above; however, that does not appear to be a factor in this case. My preference for one table over its twin is sentimental in the sense that it has a meaning imparted to it by a particular history; perhaps it was given to me, say, by my father. Likewise, my preference for my nieces over my twin's has to do with the differences in our respective histories; I have known, loved, and interacted with *the former* since their birth – and that is enough to explain my preference for them. (Of course, there is also a qualitative difference associated with my nieces and their twins if CSD is correct – they are different persons in virtue of being different subjects.)

But my preference for my continuation as a mental subject over my twin has nothing to do with our respective histories: I might prefer my mother and father to my twin's for that reason, but not the continuation of my own existence *qua* subject. That preference is purely a matter of my being that conscious subject;

I do not prefer my own survival *qua* subject because of any special attachment to my mother or father – or any other element of my history. Nor do I have a special preference that it is my birth to this particular person that brings me into existence; like many other persons, I sometimes entertain the wish that I had been born to someone else – or, more precisely, that some other body bring me into existence as a subject of experience. If there is no historical difference that explains the rationality of this preference, then any non-reductive physicalist theory presupposing the mind–body supervenience thesis would appear to be false since there are no neurophysiological differences that would explain it (or, for that matter, environmental differences if this is supposed to make some kind of difference independent of its impact on neurophysiology).

Whether or not there is a difference in *qualia*, we seem to have a mental difference that is theoretically significant for a theory of mind without any possible adequate physicalist explanation. That I care so much more about my survival as a conscious subject than my twin's suggests that my being associated with one body rather than another requires explanation by a theory of mind. But physicalism seems to lack resources to give the explanation: historical considerations are explanatorily irrelevant and there are no neurophysiological differences between me and my twin. So, whether or not the difference between me and my twin involves a difference in *qualia*, it is a theoretically significant one for which non-reductive physicalism lacks resources to explain given its commitment to mind-body supervenience. Again, the thought-experiment seems to force a choice between CSD and some intuitively implausible – at least, from my vantage point – form of reductive physicalism.

One can, of course, simply deny all of the underlying intuitions that this argument relies upon, including intuitions about the existence of a conscious self and the intuitions about survival of one's physical death, simply to preserve a commitment to non-reductive physicalism. This requires denying intuitions that have been very common among laypersons and philosophers, dualists, and physicalists alike (both Thomas Nagel's statement of the problem and the Strawson quotation in note 3 presuppose or express such intuitions), but it would be better, on my view, to acknowledge, at the very least, that the thought-experiment raises some important questions about the character of conscious subjectivity that require philosophical examination and not just a digging in of one's heels. If the thought-experiment does not show the plausibility of CSD, then it surely raises deep issues about the character of conscious experience.

Three more objections and replies

The issue is just a trivial fact about identity

One line of response claims that the issue described by Nagel poses no real philosophical difficulty because it requires nothing more than an explanation of

why I am identical with myself and am not identical with you. Thus conceived, the problem of explaining conscious subjects is no more a problem than the problem of explaining why any particular thing is what it is and not something else.

This response is problematic for a couple of reasons. First, interpreted this way, the problem is simply to explain why the following formula is necessarily true: $(\forall x)((x=x) \ \& \ (\forall y)(x \neq y \rightarrow x \neq y))$. If, on this line of interpretation, this is a problem, it is a problem that afflicts every theory – and not just a physicalist theory of mind.

But the demand for an explanation of how a particular collection of atomic, subatomic, or molecular material arranged body-wise gives rise to *me qua* conscious subject is not a demand for an explanation of the trivial claim that I am me and not someone else. It should be clear that the claim that every aspect of mentality can be explained in terms of the causal properties of physical entities, states, and processes is a substantive theoretical claim not implied by the logical truth that I am identical with myself and no one else. There is no obvious logical truth that would explain why *I* happen to be the distinct conscious subject causally associated with the operations of the electrical activity in the brain of my mother's first-born son.

In contrast, the claim that I am identical with a particular conscious subject is a claim that is, if not a logical truth, a metaphysical truth. I can conceive (or seem to be able to conceive) of the subject of my conscious experience as associated with content that is furnished by other brains in other bodies. I cannot even begin to make sense of the idea that I might be a different conscious subject: as far as I can tell, it is a brute metaphysical fact that I am not any other subject than the one that happens to exist in the body of the first-born son of my mother – and could not be any other.

Indeed, this intuition explains why Nagel's concern states a genuine problem. It assumes not only that the conscious subject existing in the body of the first-born son of my mother did not exist in any other body, but also that it is not the case that some other conscious subject pre-existed the one to whom my mother gave birth in 1957 and that one was also me. If Nagel fails to state a real problem, it won't be because he makes the mistake described by this objection.

It is trivially true that you are the subject of your body's experience

A second line of response claims that it is trivially true that one of these streams is *yours* because it is brought about by physical stimulation of *your* body, and the other is *your twin's* because it is brought about by physical stimulation of *your twin's* body. On this line of response, physicalism is consistent with this result because there is no other possible outcome from the standpoint of physicalism. By definition, for all conscious subjects A, the qualitative stream associated with A's body is experienced by A. Thus, it is trivially true that the

stream associated with your body is yours and the stream associated with your twin's body is your twin's.

Such reasoning provides a simple basis for explaining the rationality of my strong preference for my survival conceived as just a conscious subject over my twin's. On this line of reasoning, then, there are no differences that need explain why I prefer my survival to my twin's; I prefer my survival simply because I am this being rather than my twin. It could not be otherwise.

This misunderstands the concern the twin-earth argument raises for physicalism. The point here is not to demand an explanation for either the claim that, for any conscious subject A, A is the subject of A's stream of consciousness or the equally trivial claim that, for any conscious subject A, A is the subject associated with A's body. I think it is fair to say that those claims are necessarily true – though the nature of the modality is not entirely clear to me.¹⁹

Rather, the point is to demand an explanation for the fact that the conscious subject that is metaphysically identified with *you* experiences the stream that is brought about by the stimulation of one of these bodies rather than the stream brought about by the stimulation of the other. In other words, the issue, as Nagel describes it above, is why one of these collections of atoms arranged into the form of a human body produces a stream of conscious content associated with a subject that is you and the other produces a stream of consciousness associated with some other subject; that is, the issue is to explain why one of these bodies – and not the other – is *yours* and the other is someone else's. Since the two bodies are physically and nomologically indistinguishable at every relevant level of description, it is completely arbitrary from the standpoint of physicalism that one of these human bodies is *yours* and the other is someone else's.

Here it might be helpful to note that we can sensibly ask other questions about the relationship between what are essentially parts of a person. I can ask, for example, why I have this right arm attached to my body, rather than some other. In the case where it is the arm with which I was born, the explanation will make reference to the facts of reproduction, along with some facts about how DNA functions to create a body in which all cells share the same genes. What I can't sensibly ask, however, is why this arm is the one that it is and not some other; if that is even a question that makes sense, then I suppose the answer will be nothing more interesting than it is a metaphysically brute fact, first noticed by Bishop Butler, about the world that for all X, X is what it is and not something else.

My strong preference for my survival as a continuing subject of consciousness over my twin's obviously has something to do with the fact that I am the former being rather than the latter being, but it is unhelpful for our purposes. My preference for that being's survival is not necessarily tied to the survival of the body that provides that being with the stream of conscious experience it instantiates. It is not clear what sort of experience a mental subject without a body could have. But if I had to choose between being annihilated as a subject

of conscious experience during the process of physical death or becoming a disembodied conscious subject upon the death of my body, I would strongly prefer the latter. That is how important my continuing survival as a conscious subject is to me: much more important to me than the continuing survival of my body as a living physical organism; being in a permanent vegetative state is no better than death, from the standpoint of rationality.

One might object to the last line of argument by claiming that we cannot *really* conceive of our continued survival as minds without our bodies – at least not in such a way that conceivability is a reliable index to what is metaphysically possible. Thus, on this line of thinking, the argument in the last paragraph is problematic insofar as it presupposes that I can conceive of my being the subject of a stream of consciousness; such survival might be epistemically, but not metaphysically, possible.

This last line of argument doesn't presuppose either the metaphysical possibility of the subject of my experience being detached from my body or even the epistemic possibility of the subject of my experience being detached from my body. It simply presupposes the truth of a conditional whose antecedent, for all we know, might be necessarily false, but would be true even if the antecedent were true: *if it were metaphysically possible for me to survive as a subject of some sort of conscious experience not intolerably bad without my body's survival, I would want to survive in that way if it were otherwise inevitable that the body would die.* Regardless of whether it is metaphysically possible for the subject of my experience to survive the death of my body, I would prefer this if it were. The same is true for a lot of such conditionals: I do not know whether it is metaphysically possible for me to have been a well-loved, well-cared-for dog, but there are some days where I believe that if it were possible for me to have been such an animal, I would prefer to be such an animal to a human being – even if the antecedent, for all I know, turned out to be true.²⁰ So no assumptions whatsoever are made about the metaphysical or epistemic possibility of such a separation of the subject of my experience and my body.

The analysis begs the question against physicalism by presupposing the existence of an objective self

There is one more objection that is worth considering, which relies on the mysterious notion of an objective self as explained by Thomas Nagel. Expressed briefly, the objection is as follows. My description of the problem and of the thought-experiment presupposes the Nagelian notion of an objective self, which illicitly assumes, on this line of reasoning, that conscious subjects are constituted by entities that are only accidentally (or contingently) connected to the particular bodies with which they are connected. For example, the objective self that is me is contingently connected with the body of the first-born son of my mother, which implies that it is metaphysically possible that that my objective self could have

been paired with some other body. This strongly suggests, if it does not entail, that objective selves are mental substances or ‘soul pellets’ that are, as a metaphysical matter, interchangeable with other bodies and can even survive the death of the bodies with which they happen, as a matter of contingent fact, to be connected.²¹ If all this is correct, then Thomas Nagel’s description of the problem, as well as my own thought-experiment, begs the question against mental-physical supervenience.

The objection mischaracterizes what Nagel has in mind by an objective self. Nagel never suggests that these selves are contingently connected with the body; in fact, he explicitly denies, and correctly, that it follows from what he has to say about the notion that objective selves are contingently connected with bodies:

The fact that I *seem* able in imagination to detach this perspectiveless or objective self from TN does not show that it is a distinct thing, or that nothing else about TN belongs to me essentially. It does not show, as may first appear, that the connection between me and TN is accidental. (*VFN*, 61–62; emphasis added)

One might be tempted to think that he is mistaken here, but he is clear in saying only that it *seems* to us that we can conceive of this self as distinct from my body. The claim that it seems to me that I can conceive of my self as distinct from my body implies nothing about whether its connection with my body is contingent, because I can seem to conceive of all sorts of things that might not really be conceivable (at least in a sense that would imply that such things are metaphysically possible). For example, I seem to be able to conceive of a stuffed animal made entirely of cotton having thoughts and feelings (as occurs when I read *Winnie the Pooh* to one of my nieces), but the existence of such a thing may not be conceivable because the idea of something sentient being made entirely of ordinary cotton may not be metaphysically coherent.

Accordingly, there is nothing here that implies that the connection between this self and this body is contingent and hence nothing here that is fairly characterized as begging the question against the mind–body supervenience thesis.

Conclusions

The possibility of the earth’s having a twin in this universe seems to call into question any non-reductivist physicalist theory relying on the mind–body supervenience theory. The fact that I am a conscious subject associated with a body distinct from a body that is identical with mine in all neurophysiological respects entails that the difference between my twin conceived as conscious subject and me as conscious subject cannot be explained in terms of neurophysiological states. If, as I have argued, this difference is one that a theory of mind is obliged to explain, then the thought-experiment seems to be a counter-example to the mind–body supervenience thesis.²²

Notes

1. Thomas Nagel *The View from Nowhere* (New York NY: Oxford University Press, 1986) [hereafter *VFN*].
2. One of the many ways in which my thought-experiment connects with the philosophy of religion is that it implies that we cannot look for an explanation for the existence of conscious subjects that is defined wholly in terms of neurophysiology or physical processes, like the evolutionary process described by Darwin. Since subjectivity does not supervene on neurophysiology, the explanation for its existence will have to include reference to non-physical entities and processes – something that itself has the property of being conscious, which points, given that this entity would have the extraordinary power of conscious entity to bring new conscious subjects into existence, in the direction of God. For examples of such arguments, see J. P. Moreland *Consciousness and the Existence of God* (London: Routledge, 2008); Richard Swinburne *Is There a God?* (New York NY: Oxford University Press, 1996), 69–94; *idem* ‘Arguments from consciousness and morality’, in *idem The Existence of God* (Oxford: Clarendon Press, 1979), 152–175; Robert Adams ‘Flavors, colors, and God’, in *idem The Virtue of Faith and Other Essays in Philosophical Theology* (New York NY: Oxford University Press, 1987), 243–262; J. L. Mackie ‘Arguments from consciousness’, in *idem The Miracle of Theism* (Oxford: Oxford University Press), 119–132.
3. J. P. Moreland and William Lane Craig *Philosophical Foundations for a Christian Worldview* (Downers Grove IL: Intervarsity Press, 2003), 238.
4. As Galen Strawson convincingly puts the point: ‘It is very natural for us to think that there is such a thing as the “self” – an inner subject of experience, a mental presence or locus of consciousness that is not the same thing as the human being considered as a whole. The sense of the self arises almost irresistibly from fundamental features of human experience and is no sense a product of “Western” culture, still less a recent product of it, as some have foolishly supposed’; Galen Strawson ‘Introduction’, in *idem* (ed.) *The Self?* (Oxford: Blackwell Publishing, 2005), vi.
5. For example, the thesis that any world indistinguishable from ours in terms of physical entities or properties is indistinguishable in terms of non-physical entities or properties seems false; there might be a world physically indistinguishable from ours with ectoplasmic spirits that play the right functional roles to count as minds. This is the ‘problem of extras’. For a discussion of the difficulties in formulating the appropriate supervenience thesis, see Andrew Melnyk ‘Physicalism’, in Stephen Stich & Ted Warfield (ed.) *Blackwell Guide to Philosophy of Mind* (Oxford: Blackwell Publishing, 2003).
6. As Jaegwon Kim puts the matter: ‘Mind–body supervenience can usefully be thought of as defining *minimal physicalism* – that is, it is a shared minimum commitment of all positions that are properly called physicalist, though it may not be all that physicalism. As is well known, there are many different ways of formulating a supervenience thesis. For present purposes ... [i]t will suffice to understand it as the claim that what happens in our mental life is wholly dependent on, and determined by, what happens with our bodily processes’; Jaegwon Kim *Physicalism, or Something Near Enough* (Princeton NJ: Princeton University Press, 2005), 13–14.
7. Frank Jackson raised this as a concern for standard statements of the physicalist supervenience thesis. See Frank Jackson *From Metaphysics to Ethics: A Defense of Conceptual Analysis* (Oxford: Clarendon Press, 1998).
8. Daniel Dennett takes the problem of explaining selves as seriously as any physicalist theorist: ‘People have selves. Do dogs? Do lobsters? If selves are anything at all, then they exist. *Now* there are selves. There was a time, thousands (or millions, or billions) of years ago, when there were none – at least none on this planet. So there has to be – as a matter of logic – a true story to be told about *how there came to be* creatures with selves. This story will have to tell – as a matter of logic – about a process (or series of processes) involving the activities or behaviors of things that do not yet *have* selves – or *are* not yet selves – but which eventually yield, as a new product, beings that are, or have selves’; Daniel Dennett *Consciousness Explained* (New York NY: Little, Brown & Co., 1992), 413–414, emphasis added.
9. Geoffrey Madell has written on related but somewhat different problems, and his arguments have, not surprisingly, been different from the argument advanced below. For example, in chapter 4 of *Mind and Materialism* he states that physicalism is totally unable to give an explanatory account of the subjective character of experience, the what-it-is-like-to-be-ness of conscious subjects, as described by Thomas Nagel. He deploys Frank Jackson’s famous argument involving the neuroscientist Mary, responding to counter-arguments to Jackson’s argument. See G. Madell *Mind and Materialism* (Edinburgh: Edinburgh

University Press, 1981). This is a related but different problem from the problem of how bodies give rise to conscious subjects, which is the one I consider here. In chapter 5, Madell is concerned with a problem more closely related to the one with which I am concerned here; as he describes it, ‘The problem is most clearly seen in relation to the *first person*. Thomas Nagel put his finger on it in his paper “Physicalism”. Let us envisage the most complete objective description of the world and everyone in it which it is possible to have, couched in the objective terminology of the physical sciences. However complete we make this description, “there remains one thing I cannot say in this fashion – namely, which of the various persons in the world I am”. No amount of information non-indexically expressed can be equivalent to the first person assertion, “I am G.M.”. How can one accommodate the existence of the *first-person perspective* in a wholly material world?’; Madell *Mind and Materialism*, 103. This is a different problem from the problem of whether physicalism can explain, compatible with the supervenience thesis, how it is that the body of the first child born to my mother could bring me into existence as a subject of consciousness. While the statement of the problem with which Madell is concerned presupposes the existence of an embodied conscious subject (that, he argues, cannot be picked out by any language that reflects materialistic assumptions), he is not concerned with the specific issue of whether an account of the existence of such subjects can be explained by neurophysiological states. I argue that it is not possible to give such an explanation, and hence that dualism is true. At the end of the day, Madell and I are philosophical allies, but we are waging different battles.

10. The point of specifying these assumptions is to make clear that one can consistently specify a logically possible world containing two planets that resemble each other so closely.
11. Thus, for example, you and your twin experience exactly similar random thermal fluctuations and your bodies respond to those fluctuations in isomorphic and hence qualitatively indistinguishable ways.
12. Since you are presumably billions of light years apart, you could not be exposed to, for example, exactly the same sun. But the sun to which you are exposed is qualitatively indistinguishable from the counterpart sun to which your twin is exposed.
13. While this is unlikely if libertarianism is true, it is possible. In any event, it is important to note that the state of affairs in which two persons always agree in their volitions is not inconsistent with libertarianism.
14. This thought-experiment seems to pose no dilemma for reductive physicalism. The eliminativist, for example, denies that ordinary folk categories referring to phenomenal experience, including the phenomenal element of being a subject, have any referent. If this is right, then this thought-experiment does no damage whatsoever to eliminativism.
15. None of this should be taken to imply that minds and bodies cannot causally interact. They undoubtedly do; my visual experience is surely caused by the neurophysiological state of my brain. It is simply to assert that my being the particular subject I am cannot be explained just in terms of neurophysiology.
16. By ‘global’ here, I mean to pick out the most general supervenience theses, applying to all objects, properties, etc. in the universe, rather than the more narrow one focusing on minds.
17. If not, then the relevant global supervenience thesis will have to be supplemented by other theses; however, these other theses will obviously be fully compatible with the basic intuition that everything is ultimately physical.
18. The conditional holds in one direction only. I might be indifferent as between strawberry and chocolate ice cream although they differ qualitatively.
19. My intuition is that the modality is conceptual. It seems reasonable to think that, as a conceptual matter, a physical object X is A’s body if and only if A is the subject associated with X. That is, it is part of the very concept of some physical object X being A’s body that A is the subject associated with X. It should be noted that the dualist intuition that a subject A could, so to speak, inhabit some other body than the one A actually inhabits is consistent with this conceptual claim. Let A be a subject, X be A’s body, and Y be some body distinct from X. The relevant dualist claim, then, is that it is possible that X is not A’s body and that Y is A’s body. This is consistent with it being necessarily true that A is the subject of A’s body; for the claim is just that it is not necessarily true that X (which is, in fact, A’s body) is A’s body.
20. Another example is the sentence ‘if square circles exist, they are unlike any other objects in the universe’. The statement is not just trivially true in the sense that the antecedent is necessarily false;

if the antecedent were true, square circles would disprove the law of non-contradiction and hence be remarkable objects.

21. I am grateful to David Chalmers for this concern.
22. I am grateful to Jaegwon Kim, David Chalmers, Jennifer Corns, David Pereplyotchik, Stephen Layman, Patrick McDonald, and Rebekah Rice for their attempt to help me with an argument that they did not find compelling.