Is Searle a Property Dualist?

Felix Schröder
## Contents

1. Biological naturalism, recounted 2

2. From non-reduction to exclusion 11

3. Biological naturalism, reconstrued 13

4. Conclusion 20

References 21
Ever since the publication of Searle’s *The Rediscovery of the Mind*, the conception of the mind therein presented, dubbed “biological naturalism” by the author, has been faced with persistent accusations of disguised property dualism, which have tended to base themselves on a perceived incompatibility between Searle’s nonreductivism and physicalism. I intend in this paper to argue that this is a mistake.

I will begin by giving a short summary of Searle’s theory as he originally presented it. This will be followed by an analysis of the claim of property dualism against him in one of its most standard iterations. In so doing, I will attempt to demonstrate that it is based on a misunderstanding of Searle’s concept of ontological irreducibility. This misunderstanding arises due to Searle treating reduction as a metaphysical relation, but irreducibility as the failure of a method used to prove or bring this relation about as opposed to its negation, something overlooked by his critics. Due to the nature of this method and the reasons for its breakdown, irreducibility turns out to be a semantical relation which does not entail anything of particular significance about the metaphysics of the relata. Finally, in order to acquit him of the charge of the charge of property dualism, I will argue that the inapplicability of the reductive method Searle discusses to consciousness does not commit him to property dualism. To this end, I will demonstrate it to be compatible with a form of identity theory that deposits consciousness firmly within the physical realm.

1. Biological naturalism, recounted

In order to explain biological naturalism, one must first describe the ontological framework in which it is based. Searle sees the world as constituted by a sort of tiered structure of physical systems, or as he puts it:

> All the big and middle-sized entities in the world such as planets, galaxies, cars, and overcoats, are made up of smaller entities that are in turn made up of yet smaller entities until finally we reach the level of molecules, themselves composed of atoms, themselves composed of subatomic particles (Searle, 1992, p.86)
In this tiered structure there are at the fundamental level only elementary particles, and the higher levels are occupied by systems with these as their constituents, systems being defined by Searle as “collections of particles where the spatiotemporal boundaries are set by causal relations” (ibid pp.86-87). The common observation that what is true of the parts need not be true of the whole and vice versa applies to these systems as well, in relation to the lower-level entities that constitute them. Those properties that the system possesses while each part does not, Searle calls system features. Many of these are simply sums of the properties of the components taken as a whole, such as the mass or shape of the system. The mass of a car is generally at the order of magnitude of 1t, but this will not be true of its individual parts. The constituents all taken together, however, exactly match the mass of the car, and so the mass of the car can be deduced by looking only at the individual constituents in isolation. Likewise, just by looking at the individual positions of the components, the shape can be deduced. However, Searle distinguishes from these a further group of system features, those he calls “causally emergent”. Examples are properties like the phases of matter, which exist only as causal interactions within the system. Solidity does not exist as a property of molecules, nor is the solidity of a system deducible from looking at the individual properties of such. It depends upon whether they attract each other with sufficient force to resist deformation, so one must instead look to the electromagnetic interactions between the molecules (ibid p.111).

According to Searle, the mind is a system-level feature of the brain, and belongs to this second category, i.e. it is causally emergent. Just as one cannot deduce the phase of a system by looking at the particles in isolation, so one cannot deduce the conscious state of the brain by looking at neurons in isolation, one must instead look at their interactions (ibid p.112).

Up until this point, Searle’s views have been comparatively uncontroversial, but having now established his basic views on how consciousness relates to the brain he makes a further claim: Consciousness is irreducible to lower-level features of the brain. Placing consciousness ontologically in the position Searle now has, as well as insisting on irreducibility, is generally
taken to be the hallmark of what is called emergentist property dualism, of which Searle has consequently been accused. For him this would be an utterly unacceptable consequence for the theory, as to him property dualism entails the causal impotence of the mind, one of the conclusions biological naturalism is specifically designed to avoid (Searle, 2002).

Property dualism is, in general, the view that all mental properties are non-physical, though I expect it would be fair to call any position on which at least one mental property is non-physical a property dualist one. The emergentist form is simply one which does not postulate non-physical properties at every level of reality, instead holding that as we move up the hierarchy and the systems and processes become larger and more complex, properties unrelated to those at the lower levels suddenly start appearing, “emerging” from the mass of properties the system has in virtue of its components. As for what it means for a property to be non-physical, views diverge on the issue but Stoljar (2015) divides these roughly into two groups. One defines a physical property as one ascribed to objects by theories in physics or else metaphysically depending in the relevant way on properties of such types. The other defines it in terms of being a requirement of a complete description of the intrinsic nature of “paradigmatic physical objects”, as well as their components, such as the material objects we see around us, or once again being in the right way metaphysically dependent on such. The first thing to note from this is that there is no serious dispute over the fact that any property which physics makes direct explanatory use of, such as mass, charge, spin etc., is physical. These are both attributed to objects by physical theory and required for complete descriptions of “paradigmatic physical objects”, physical theory being the framework for such description. The only other significant point to mention is the metaphysical dependence relation. While it is not clear what exactly this relation is, what is clear is that there are some properties that are not directly attributed to objects by physics, but that nonetheless do not commit one to property dualism. Take Searle’s example of solidity. Nowhere in physics will you find this as a property directly attributed to objects, yet due to the way that it is based in properties that are, it would be a very implausible view that called it non-physical. From this I derive the following, slightly unwieldy but quite effective, conditional: If a property is directly attributed to objects in physical theory, or stands in a relation
to such metaphysically equivalent to that of solidity, then it is physical. This will suffice to show that Searle’s is entitled to consider his mental properties physical.

On to the notion of reduction. Searle identifies ambiguity in the term between several reductive patterns, each of which with a corresponding relation of the form “nothing-but”, where A reducing B consists in B being nothing but A. Of those he identifies, three will be relevant here.

First of those with which we need concern ourselves, since I foresee use for it in later discussions, is what he calls logical or definitional reduction. This is characterised by Searle as a relation between linguistic entities such as words or sentences which holds when those referring to, or in the case of sentences including terms referring to, one type of entity can be translated, with no semantic loss, into sentences about making reference to another (Searle, 1992, p.114). Searle gives the example of reducing the language of numbers to that of sets, or averages to the samples whose averages they are. The reason for including this somewhat arbitrary requirement that the reducing and reduced group of linguistic entities each make reference to a single type of further entity is that this way the reduction also achieves a reduction between the referent entities. Naturally, perfect translatability is possible between any linguistic entities, whether they make reference to anything or not, depending only on their semantics matching. However, given that all sentences about one type of entity can be translated into sentences about another, the former will end up nothing but the latter, much like numbers are supposed to be nothing but sets, or so Searle claims.

This ontological form of reduction, which is of the greatest significance to biological naturalism, is presented as “the form in which objects of certain types can be shown to consist in nothing but objects of other types” (ibid, p.113). We will assume that the relation in question is that of tokens of one type being “nothing but” tokens of another in the sense of being numerically identical to such. This is possible between objects, such as chairs and collections of molecules, or between

---

1 What I mean by that we will assume it to be so is that while it is clear that Searle thinks that reduction obtains in cases of identity between the tokens, it is not clear that it obtains only in such cases. In the case of definitional reduction he speaks as though “nothing-but” is a relation that can also hold between one object on the one hand and many on the other, such as a sentence about an average being so related to all the individual
properties, such as heat and mean molecular kinetic energy. While the nothing-but relation is simply numerical identity in this case, the process of reaching an ontological reduction is not as simple as standard proof of numerical identity. In fact, the process relates more closely to the types than the tokens. It has to be shown that the tokens of the one type are always numerically identical to tokens of the other, which amounts to showing that the criteria for being a token of the one implies fulfilling the criteria of the other.

In cases like consciousness, where we have an observed phenomenon as a property of physical objects, Searle appears to think that there is only one method available to accomplish this (ibid, p.121), the inapplicability of which consequently implies ontological irreducibility: First, we start out with a term referring to the property, defined in terms of the epistemic basis for our belief in it. For example, colour terms are defined in terms of conscious colour experience and solidity is defined in terms of both its effects on our sense of touch and the fact that solid objects are to some degree impenetrable to other objects. In other words, some property terms of this kind are purely defined in terms of experience while others also make use of surface-level causal relations to objects other than us. Then, when we then discover that colour, as initially defined in terms of how it appears to us, is caused by some underlying structure, ontological reduction is accomplished by redefining the terms relating to colour with reference to the underlying causality, treating the “real colour” from that point on as distinct from the experience (ibid, p.119). Upon redefinition, the “nothing-but” relation is no more than a trivial logical consequence of the new definition.

The question here is whether this is done for epistemic or ontological reasons. Simply put, was colour “nothing but” the underlying causal property from the start, or did “red”, for example, only start referring to this property after the redefinition. Did we redefine to achieve ontological reduction between the referent of “red” and the causal basis or just because that way, as it is then a matter of definition, no demonstration is needed? At one point Searle speaks of the initial pre-theoretical definition of “real red” as being of the form “what looks red under normal

sentences about the sample members. Since he is unclear on this point and it has no significant implications here, we will assume that like identity it is purely a one-to-one relation.
conditions to normal observers” (ibid, p.115). He also mentions that “real heat” would originally have included the feeling of heat as part of the property, but that this was excluded by the redefinition and ceased to be treated as part of the property itself, becoming an effect of it instead. Searle does not go into great detail about what this “experience being part of the property itself” amounts to, but to me, there seem to be two possible interpretations of it.

According to the first, what we refer to as “having the experiential quality of red when viewed by normal observers under normal conditions” or “having the experiential quality of heat” were originally treated as properties of objects. These are of course not really intrinsic to the objects we call “red” or “hot”, these are really qualities of our conscious experience, so this would involve projecting our experience on the world. This way, the experience is ontologically part of the property, the experience at least in part constitutes the property, even if in the case of red at least\(^2\) the normalcy conditions mean that just having the experiential quality alone is not enough to instantiate the property. In this case, it seems the redefinition really does bring about the “nothing-but” relation. Before redefinition the property picked out is the experiential quality, mistakenly attributed to the object. After redefinition, it is the cause of said quality. Since the former is not a property of the object at all, clearly it cannot be identical to the latter.

According to the second, the definitions before reduction were descriptions. “Looking red to normal observers under normal conditions” is not in this case treated as specifying a complete property of its own in the same way, but analysed as a description like “whatever property makes objects look red to normal observers under normal conditions”. Whatever the current definition of red is, for example, we know that we can pick out that property with a description like “whatever property, traffic lights having which makes cars stop before them”\(^3\) or somesuch. If we wanted, then, we could use this as a definition of red, though it would be arbitrary and less than informative. Thus, on this interpretation, this is what Searle is saying that we originally did, without knowing what the property picked out really was. The experiential quality was if so

\(^2\) Though it is natural to assume the same is true of temperatures, we know for example that in cases of extreme cold people may feel warm and vice versa, so to rule out these cases there would probably be normalcy conditions just like in the case of colour.

\(^3\) Well, if we wanted something less transient, this example being entirely dependent upon the laws of current civilizations, we could specify a specific car at a specific time, since what has already occurred will not change depending on future laws in this way.
never a metaphysical component of the property itself, it was simply an epistemic basis for it built into the definition. The sense in which it was part of the property is that the property was by definition its cause, meaning nothing could fail to cause the experience and still have the property. If so, it is plausible that the property picked out was always the same, and that reduction did no more than replace a narrow description of a causal property’s special cases with a more general one. The cause of the experience of seeing red also causes many other things, so the original definition is changed for being excessively narrow and uninformative, not in order to pick out something new. That this logically guarantees ontological reduction is no more than a side effect.

To attempt to settle the issue, let us return to the ontological reduction process itself. Searle argues that our discovery that heat as initially defined is caused by some features of the microstructure only gets us as far as a simpler, third reductive pattern he calls causal reduction; “a relation between any two types of things that can have causal powers, where the existence and a fortiori the causal powers of the reduced entity are shown to be entirely explainable in terms of the causal powers of the reducing entity” (ibid, p.114). Having achieved causal reduction, since we are interested in causally mapping the world, when we discover that we have been speaking about effects rather than causes all this time we redefine our terms to align with our interests. This strikes me as conforming better with the view that the pre-theoretical definition was a description, since if our initial definition was a projection of our experience onto the object, it becomes difficult to see how the causal reduction relation could hold between it and anything in outside of us. The experience turns out to be a process in our minds, so whatever causal powers it has are the causal powers of that process. If the initial definition did not pick out something causally active in the object responsible for the experience and the reduction proceeds as Searle describes it it becomes difficult to see why heat was not reduced to mental processes rather than their causes.

As we will see in the next section, this is of vital importance to the fate of biological naturalism. If biological naturalism includes the claim that consciousness is ontologically distinct from, and
cannot be reductively redefined to be identical with, any state or other property that can be referred to in the language of physics and biology, then property dualism inevitably follows.

The salient point here is that in denying the reducibility of the mind, Searle is arguing only that the relevant kind of pragmatically motivated redefinition cannot take place. It is the failure of a method, not a relation. In general when dealing with observed phenomena in the world we want to know what is causing the observations and how in the greatest possible detail, and go about defining our words to accomplish this, which is the end towards which ontological reduction serves as a means. With respect to consciousness this is supposed to be impossible, since the language of conscious experience is not defined in such a way as to allow us to get rid of it by discovering the underlying causal reality even given a causal reduction in terms of brain processes or particle interactions or what have you (ibid, pp.120-121). We generally redefine so as not to make reference to anything other than what interests us most about the property, which in cases like colour is not how it relates to us through our subjective experience but what it is really there in the object, in that case the cause of the experience in question. However, in the case of consciousness what we are most interested in, in fact the whole reason terms for describing consciousness were created in the first place, is to represent the experience and not its causal basis (ibid, p.121).

Searle is here asserting only a meaning relation between the language of the natural sciences and the language of consciousness, namely that the latter cannot be redefined into the former without losing meaning that is of interest to us. What this is very specifically not saying is that there is any “real” separation between the two, something Searle is very insistent upon.

Consciousness fails to be reducible, not because of some mysterious feature, but simply because by definition it falls outside the pattern of reduction that we have chosen to use for pragmatic reasons. [...] The contrast between reducibility of heat, color, solidity, etc., on the one hand, and the irreducibility of consciousness on the other hand, does not
reflect any distinction in the structure of reality, but a distinction in our definitional practices. (ibid, p.123)

The failure of the reduction relation, which is a metaphysical relation rather than a semantical one, would quite plausibly constitute such a distinction in the structure of reality. Hence, this conforms far better with the view that the all irreducibility amounts to for Searle is the breakdown of a reductive process. If reduction is the “nothing-but” relation, reducibility is the relation that holds when the terms for the relata are so defined as to allow reductive redefinition with the goal of bringing reduction about. Another telling passage to that effect from his later writings is the following:

The sense in which, though causally reducible, [consciousness] is ontologically irreducible, is that a complete description of the third person objective features of the brain would not be a description of its first person subjective features. (Searle, 2002)

Now if what he meant by consciousness being irreducible were that the ontological “nothing-but” relation fails to hold, this would simply be false. The sense in which consciousness is irreducible would if so be that this relation fails to hold, wholly unrelated to any relation between descriptions, since the latter is an entirely linguistic matter. It fits perfectly, however, with irreducibility being a relation between two objects holding when the terms for the relata do not admit of reductive redefinition in terms of one another. The descriptions fail in this case to be equivalent, not because the states they describe differ metaphysically, but because the descriptions themselves differ significantly in semantics.

In summary, Searle’s actual contention is neither concerned with whether this relation obtains, nor with whether there is as of this moment definitional reducibility between mental and for example biological terms. The former as we will see does indeed imply property dualism, while on the other hand the latter could be granted even by a reductivist given Searle’s analysis of reduction. The issue is whether, given the fact that there is not definitional reduction, it could be
about through reductive redefinition. Searle denies this only because it would be a loss pragmatically from the human perspective, this implying a breakdown of the reductive method as he understands it. We would become unable to convey semantic content of interest to us, because this content is of a kind that cannot be reconstructed or compensated for in a satisfactory way by language defined with reference to publicly observable phenomena. That is all there is to the irreducibility claim.

2. From non-reduction to exclusion

Now that we know what biological naturalism is, as well as what it means to say that its irreducibility claim entails property dualism, we can proceed to see why it has been believed to hold true. Jaegwon Kim writes:

Are these higher-level physical features of the brain (a.k.a. mental properties) reducible to, or reductively identifiable with, the lower-level properties [...]? Searle's answer, like the property dualist's, is a forceful no. But it is precisely this negative answer that defines property dualism. (Kim, 1995, p.192)

Kim, like most philosophers of mind it is fair to add, treats reduction in quite a different way from Searle. To him, there are two aspects of reduction to be separated; the reduction itself and reductive explanation. He treats reduction and reducibility purely as synonymous terms for the metaphysical relation, and it is for all intents and purposes the same “nothing-but” that Searle makes use of. Also like Searle, he is slightly unclear on whether this relation consists in or simply includes the relation of numerical identity, however differing in explicitly expressing uncertainty about the issue. See for example:

If pain = [neural state] N1, there is no pain over and above N1; and if mental states are identical with brain states, there are no mental states over and above brain states. This is an open-and-shut affair if anything in philosophy ever is: Identities do reduce. For
reduction nothing works as magically as identities, and it may well be that identities of some sort are required for any genuine reduction. (Kim, 2008, p.100)

As for reductive explanation, Kim speaks of this simply as explaining one group of phenomena in terms of another (ibid., p.94), a purely epistemic matter which might be achieved even without the reduction relation, or at least is defined in a way compatible with this.

His misunderstanding of Searle probably comes primarily from the way he treats reduction, rather than reductive explanation, so I will not dwell on the latter. If Searle were discussing ontological reducibility as equivalent to the reduction relation and denying that this relation holds between mental entities and any of the lower-level entities making up the populations of biology and physics, this would indeed leave him vulnerable to Kim’s subsequent overdetermination argument (Kim, 1995, pp.192-194), serving so to speak to causally crowd his mental properties out of physics.

The overdetermination argument in question runs as follows: If mental properties are distinct higher level properties, meaning they are irreducible to and therefore ontologically additional to the properties of the brain we know, then we can look at their causal interactions. For any event we might think is caused consciously, say me lifting my arm, we know of a non-mental physical state of the brain, characterised entirely in terms of particles and their interactions, which causes that same event. Thus, given that the mental state is really ontologically something in addition to the that brain state, we now have two causes, each sufficient, for that one event. This renders the mental causes superfluous and subject to Kim’s exclusion argument (Kim, 2010, p.216). Searle too agrees that if mental causation implies overdetermination then epiphenomenalism, the view that mental events do not cause physical events at all, follows (Searle, 2002). Unlike when treating ontological irreducibility as a semantic fact, however, when treating it as genuinely ontological, causal reducibility will not save him from this. He may assert that the causal powers of the mental state can be entirely explained in terms of the causal powers of the brain state, but

---

4 See section 3 for how it accomplishes this when correctly interpreting Searle.
there are nonetheless two ontologically distinct states engaging in causation here and unless we can somehow get that to add up to a single causal event that means two causes. Given a single causal event and two entities doing the causing, surely each would have to contribute a part rather than the whole of the causal event, but that is not the case here. Each cause is to be sufficient, which leaves us with incoherence should we appeal to causal reduction.

Consider Searle’s (2002) statement to the effect that “the mark of empirical reality is the possession of cause and effect relations”. Taking into account that the physical world is essentially defined as the empirical world, he would surely once more agree with Kim and concede that this epiphenomenalism implies property dualism. In other words, Kim’s argument would be quite decisive if Searle did indeed hold the view Kim probably believes that he does, in one stroke proving both that biological naturalism is property dualism disguised and that it is no better suited to preserving mental causation than other forms of this view.

3. Biological naturalism, reconstrued

Having now shown that this standard assumption used to show that biological naturalism is property dualism is based on a mistake, the final step is to show that correctly understood it does not contradict physicalism. To this end, I will sketch the view that I believe that Searle is grasping for, though this may, to paraphrase a famous introduction to Kripke’s *Wittgenstein on Rules and Private Language*, be straying into the territory of “expounding neither my view nor Searle’s, but rather Searle’s argument as it struck me”. No matter, regardless of whether this is the view that Searle counts as his own, the mere fact that it is compatible with biological naturalism is sufficient to prove that biological naturalism is not committed to property dualism.

Biological naturalism makes a number of claims about consciousness. That it is in a sense emergent, but only in a way compatible with its causal reducibility, meaning that no causally active property can be “squirted out” by a complex system and develop “a life of its own”

---

5 Incidentally, this means that interpretations of Searle like Corcoran’s (2001), where mental properties are physical/biological yet stand ontologically apart from other physical properties in this way, being therefore causally impotent on pain of overdetermination, are not compatible with Searle’s notion of physicality and hence cannot be correct. If Searle is a property dualist he is an epiphenomenalist, and if he is an epiphenomenalist he is a property dualist.
(Searle, 1992, p.112). That the causality of the universe can be exhaustively captured by a physical description of the bottom level of reality (ibid, p.87), meaning that insofar as anything is causal at a higher level it is, in the end, in the sense of aggregating the causality of elementary particles. That the language we use to describe it has semantic content of interest to us not to be found in the language used to speak of the reduction candidates we have for it, making it irreducible. That it is causal. These are the claims that must be reconciled with physicalism if my goal is to be accomplished.

The easiest way to make consciousness causally emergent and causally reducible would be psychoneural identity of some kind. We know what events we want mental causation involved in, and we know that the physical causes of these are in the brain. Searle even goes as far as to affirm something of the sort explicitly, claiming repeatedly that consciousness is an ordinary biological/physical feature of the brain like any other. However, Searle also expends great effort seemingly rejecting in the strongest terms every possible way to render this coherent.

First he rejects type-type identity, according to which for each type of mental state(pain, joy etc.) there is a type of physical state of the brain(neuron firings and the like), such that each token of the former is identical to a token of the latter. He does this because he perceives it as trying to “get rid of the mind”(ibid, p.37), in essence trying to redefine it out of existence by the ontological reduction process outlined earlier. “Either materialism of the identity variety leaves out the mind or it does not; if it does, it is false; if it does not, it is not materialism.” (ibid, 37) He also expresses some sympathy for the multiple realizability objection, i.e. he does not want pain to be identical to neuron firings of the kind present in human brains since then without those neurons one could not feel pain. Each mental state should be possible to realize in quite varied physical systems.

He goes on to reject the other option, token-token identity, according to which there are no requirements of type correspondence like this, but there is for every mental state token a brain state token to which it is identical. This rejection is on the basis that the only way he knows of to
specify which mental state tokens are identical to which brain state tokens without type identity is through functionalism (ibid, p.40), which falls victim to his famous Chinese Room argument (ibid, p.45). The kind of functionalism he discusses treats the whole conscious system, like for example the entire brain, as a sort of black box. It then defines consciousness as this black box producing from the same inputs the same outputs a human would, in other words responding to external stimuli and behaving like a human. Searle does not think that this is enough; the causal relations to and from the mind is not what defines it, he thinks, it is the qualitative aspect, and behaving indistinguishably from the outside does not prove the right kind of process for this is going on inside.

However, there is a further form of identity theory, one which neither attempts to get rid of useful semantics in favour of those of the language of physics or biology through reduction, like the type-type identity Searle discusses, nor entails that passing a Turing test\(^6\) implies consciousness by definition, as the form of functionalism he rejects would have it. Recall that the epistemic bases we had for belief in various properties were always the conscious experiences they caused in us, on the basis of which we defined the terms for the properties. Since we couldn’t tell anything about the properties other than what they caused in us, we ended up talking about the properties as the causes of our experiences, whatever those might be. Perhaps this never changed. We redefined our terms when we discovered more general ways to specify the causes, but do the current definitions not still rely, while perhaps in more roundabout ways, on the ways in which these properties can bring about experiences in us? On Searle’s analysis the reduced definitions are invariably the most general causal specifications we know, and we only ever know about what something is able to cause is to the extent that we receive evidence of this in the form of conscious experience. The only exception is consciousness itself, which is its own epistemic basis and where “the appearance is the reality” (ibid, p.122). Uniquely, we have direct epistemic access to it unmediated by any causal interactions.

\(^6\) Which is possible iff the external behaviour is indistinguishable from that of a human by the definition of such a test.
So we know that consciousness, and only consciousness, is known to us as a property our epistemic basis for belief in which which is not in any way causal. That is not all, however. On Searle’s view, our only justification for belief in causality comes from our experience of mental causation (Searle, 2011, 13:14), and we then extrapolate from what we do when we cause and are affected mentally to interactions between objects in the world. Searle shares much of his perspective on causes with Hume, meaning that not only our belief but our very conception of causation is acquired this way. Hence, given that we then only have a conception of what a cause is insofar as it is like the mental causes we experience, if we know that anything is causal we know that mental states are. These twin facts; that we know conscious states as conscious states and not as causal, and that we also know that in addition to their non-causal first-person aspects they are also causal if there are causal entities at all, might allow us to move further. If the very source of our conception of causality lies in mental causes that at the very least have other aspects in addition to their causal powers, when extrapolating to the world we seem at least to have no reason to assume that other causal properties do not. We know, at a bare minimum, that there are some causal properties that are causal yet not exhaustively specified by their causality. Of course, this leaves open the possibility that there might be properties that really do have nothing to them but causality, and without the direct experiential access we have to consciousness justifying the belief that properties outside us are not like that rather than simply suspending belief may be difficult. It is not impossible, however, and as we shall see this direct experience can be leveraged to make plausible that at least some of the fundamental physical properties are not exclusively causal, opening the route to physicality for consciousness.

Since assuming Searle’s analysis of causation is correct we know that those events we experience as consciously caused really are consciously caused, when we discover in neuroscience or physics another state which causes the same events at the same level of reality, the only path open to us is to consider this identical to the conscious state. As we saw in the last section, the alternative is overdetermination and epiphenomenalism. The physical/biological state can be decomposed into the causal properties of elementary particles and their interactions, hence so can the conscious state. This is how we use conscious experience to justify belief in
similar aspects in the properties of fundamental physics. Causally active brain states have for their components the causal properties of elementary particles. As I mentioned at the beginning of this section, Searle holds that there is nothing to the causal interactions of higher-level states other than the causal interactions of their components. If the causally active conscious state is identical to the brain state, then, it must also have for its components those same properties of elementary particles. Where the non-causal aspects of the conscious state would come from if not from its parts is difficult to grasp, so the reasonable conclusion here is that whatever fundamental physical properties are involved in causing what we cause consciously have more to them than their causal aspects.

David Chalmers gives a good summary of how consciousness can be approached on this basis:

> It can reasonably be argued that physical concepts have their reference fixed by some dispositional role, but refer to an underlying categorical property. If so, their primary intensions pick out whatever plays a certain role in the world (irrespective of its categorical nature), while their secondary intensions pick out certain instances of a categorical property (irrespective of its role). (Chalmers, 2002, p.197)

This “reasonable argument” is that the dispositional definitions of physical properties (the primary intensions), i.e. the definitions in terms of causality, work like descriptions of the form “whatever property displays such-and-such dispositions to cause and be affected”, and that this description picks out another property (the meaning of the term for which is the secondary intension). Chalmers calls this property categorical as opposed to dispositional in that it does not consist in having any particular dispositions. As we saw, the only property like this we know of according to Searle is consciousness. One might be tempted, then, as Chalmers is, to say something like that consciousness is the only categorical property and that every physical object has it. Chalmers’ favoured version of this positions is what is called panprototpsychism, according to which while not everything has consciousness like ours, there is so to speak something it is like to be any kind of physical object, including an elementary particle. This is a
view Searle, quite rightly in my opinion, rejects as “absurd” (Searle, 1997). That assumption is as unnecessary as it is implausible; not only does a thermostat contain nothing even remotely equivalent to the processes we know to correspond with experience in the case of the brain, the categorical properties can fill the explanatory roles we need without us guessing at their natures at all.

We need not and probably cannot consider what these properties of elementary particles are categorically if they exist. All we need know is that, since we have not the first idea of their categorical nature, for all we know they are capable of bringing about conscious experience when interacting with one another. What we do know is that they will be the properties into which we can decompose the physically described state, meaning things like mass, spin, charge and others. The properties directly ascribed to objects by physics, in other words, which are as established earlier uncontroversially physical. Since we are here assuming that these too have things that could if knowable to us be said about them other than what is publicly observable, there is nothing in its irreducibility that makes consciousness metaphysically different from other properties causally emerging from them. Consciousness is an aggregate of causal interactions involving basic physical properties, at least some of which categorical properties with causal dispositions. In other words, it consists in large numbers of interactions between elementary particles across the system whose property it is, and these interactions are caused by those properties of said particles that are causally active. The same is true of solidity, for example, another example of Searle’s causally emergent properties. Regardless of whether every fundamental physical property has more to it than causality, charge, given that electromagnetic interactions are certainly among those involved in mental causation, following this reasoning does. This is also the property causing most the interactions making up solidity. In fact, this is a property of every single physical object we knowingly encounter in our daily lives, meaning this hidden aspect would be quite ubiquitous indeed.

Semantically speaking, attributing a physical property to something as “whatever property has such-and-such causal dispositions” and directly as the categorical property this picks out are
clearly different. This accounts for the semantics lost if one were to redefine the language of consciousness from doing the latter to the former, explaining the irreducibility. This is also the case for solidity, the only difference being that we lack access to the categorical nature in that case and cannot define our terms with reference to it as a result. If the categorical nature of solidity were such as to be knowable, whether like in our case by the solid object itself or by anything else, any term defined on the basis of this property would be just as closed to reductive redefinition as the language of consciousness is to us. There seemed to us to be a metaphysical difference because the property of solidity appeared to be entirely publicly observable while consciousness was not. Consciousness apparently related to elementary particles in a very strange way, where just by building a complex brain out of these entirely publicly observable and knowable building blocks we could get something fundamentally different, something that was not publicly observable or knowable. As soon as the presence of hidden properties like these is generalised to cover every object in the world, the metaphysical strangeness disappears.

With this, finally, I am ready to perform that deduction the attentive reader will have noticed me laying the groundworks for in section 1:

P1. If a property is directly attributed to objects in physical theory, or stands in a relation to such metaphysically equivalent to that of solidity, then it is physical.

P2. Biological naturalism is compatible with all mental properties standing in relations to properties directly attributed to objects in physical theory metaphysically equivalent to that of solidity.

C. Biological naturalism is compatible with all mental properties being physical.

This conclusion is precisely the negation of biological naturalism implying property dualism, *quod erat demonstrandum*. 
4. Conclusion

I have here attempted to show that biological naturalism commits itself only to a semantical thesis about the relation between physical and mental language, and that this thesis is fully compatible with all objects and properties in the world being physical. The kind of view I have sketched here based on the tenets of biological naturalism is quite well suited to solving many of the traditional problems in the philosophy of mind, which is why I would urge that biological naturalism not be discarded just yet as some ridiculous disguise for property dualism due only to the lack of clarity in its presentation. This is not to say that it is without difficulties, of course. For a reader who found it convincing so far, I have an interesting problem, one which appears to plague this view more or less uniquely, to ponder as a first step toward further developing it: If the first-person private aspect of consciousness fundamentally comes from the private categorical properties of elementary particles, how do we make sense of the fact that every single elementary particle of the brain is replaced many times over during a lifetime while consciousness persists? Presumably we could perfectly replicate a human body and all its categorical properties without replicating the exact consciousness within, in the sense that the original conscious first-person perspective would not in that situation become filled with the experience sets of both individuals. Thus, the question is what distinguishes the process of creating a new person with equivalent categorical properties from replacing the existing particles with categorically equivalent ones in virtue of which the latter preserves consciousness while the former does not.
References


