

Illusionism and A Posteriori Physicalism; No Fact of the Matter

By Christopher Devlin Brown and David Papineau

[forthcoming in the *Journal of Consciousness Studies*]

Abstract: Illusionists and a posteriori physicalists agree entirely on the metaphysical nature of reality—that all concrete entities are composed of fundamental physical entities. Despite this basic agreement on metaphysics, illusionists hold that phenomenal consciousness does not exist, whereas a posteriori physicalists hold that it does. One explanation of this disagreement would be that either the illusionists have too demanding a view about what consciousness requires, or the a posteriori physicalists have too tolerant a view. However, we will argue that this divergence of opinion is merely an upshot of the semantic indeterminacy of the term ‘conscious’ and its cognates. We shall back up this diagnosis by showing how semantic indeterminacy of the kind in question is a pervasive feature of language. By illustrating this pattern with a range of historical examples, we shall show how the dispute between the illusionists and their a posteriori physicalist opponents is one instance of a common kind of terminological imprecision. The disagreement between the illusionists and the a posteriori physicalists is thus not substantial. In effect, the two sides differ only about how to make an indeterminate term precise. The moral is that they should stop looking for arguments designed to settle the dispute in their favour.

Keywords: a posteriori physicalism; illusionism; consciousness

1 Introduction

Illusionists about consciousness say there are no conscious states.

In their view, consciousness requires features that cannot be physically realized. At the same time, they uphold the physicalist doctrine that concrete reality contains nothing that is not physical. So they conclude that there are no conscious states.

A posteriori physicalists reject this elimination of consciousness.

They agree that concrete reality contains nothing that is not physical. But they don't accept that conscious states call for features that cannot be physically realized. So they see no reason to deny that conscious states exist.

What are we to make of this? The illusionists and the a posteriori physicalists agree entirely about the contents of concrete reality. Their positive metaphysical visions are identical. Yet they disagree on something apparently fundamental—whether humans and other animals are conscious. How is this possible?

One possibility is that one side is wrong about the requirements for consciousness. Either the illusionists have too demanding a view about what consciousness requires, or the a posteriori physicalists have too tolerant a view. If that is right, then we might hope that further investigation and analysis will show which side is right.

In this paper, however, we shall offer a different diagnosis. We shall argue that the divergence is merely an upshot of the semantic indeterminacy of the term 'conscious' and its cognates. The difference between the illusionists and the a posteriori physicalists is thus not substantial. In effect, they disagree only about how to make an indeterminate term precise. We shall show that this diagnosis fits well with the apparent intractability of the dispute between the two sides.

We shall also show that this kind of indeterminacy is by no means unusual.¹ Similar cases are common, not just in philosophy, but also in science and everyday thought. This is not an accident. Established background views will often militate against refining the meanings of our terms

¹ Niikawa 2021 also argues that illusionists and physicalists diverge only in attaching different meanings to the term 'conscious'. Our aim here is to further defend this analysis by showing how the possibility of such divergence is a natural consequence of the way contents attach to the general run of meaningful terms. Some other papers which are in the same general ballpark as ours are Lewis 1995—who says that if our qualia concept requires that we know the essential nature of qualia, then materialists should not believe that qualia exist, but since part of the concept refers descriptively to a functional role, materialists can accept the existence of qualia in some sense—Johnson 1992—who argues that it is a matter of mere convention whether we accept that colours exist—and Braddon-Mitchell 2003—who argues that we ought to apply a two-dimensional semantics to our phenomenal terms, such that whether 'qualia' necessarily refers to something non-physical or not depends on both (i) whether we are using the primary or secondary intension of the term, and (ii) whether actual-world qualia are physical or not. Though there are important points of agreement we have with these authors, we differ from Lewis in our rejection of a priori analytic functionalism, from Johnson in our acceptance of an analytic-synthetic distinction, and from Braddon-Mitchell in our rejection of a two-dimensional semantics for phenomenal terms and concepts. As the paper progresses, the importance of these differences will become apparent.

beyond a certain point, since they imply that the terms would pick out the same actual referents however they were refined. Sometimes, however, newly discovered facts show those background views to be deficient, and then the indeterminacy is exposed, because the new facts force a choice between eliminating or reducing the category at issue. Previous usage leaves both options open, even for those who agree about all the substantial facts. In such cases, the divergence is not substantial. It merely reflects different choices about how best to refine an indeterminate term when such refinement proves necessary.

We shall illustrate this pattern with a range of examples, and show how the dispute between the illusionists and their a posteriori physicalist opponents is one more instance of the phenomenon.

2 Illusionism and a posteriori physicalism

Allow us to begin with a closer look at illusionism and a posteriori physicalism. Both views emerged as physicalist answers to the ‘hard problem’ of consciousness (Chalmers 1995), which is the problem of explaining the relationship between subjective experience—sometimes referred to with such terms as ‘phenomenal consciousness’, ‘what-it-is-like-ness’, or ‘qualia’—and the physical brain. The hard problem is hard because it does not seem possible to conceptually infer knowledge of subjective experience from any purely physical account of the brain’s neural, functional or informational processes.

Though both illusionists and a posteriori physicalists agree that physicalism is true—and so there is nothing more to human beings beyond our physical properties and the properties which depend on them with metaphysical necessity, like functional properties—they disagree on the correct solution to the hard problem. Illusionists think that the hard problem is ultimately unsolvable. This is because they think that the meaning of the term ‘consciousness’ (and its cognates) is incompatible with any physical realization. In their view, our concept of consciousness² itself imposes requirements on conscious states that are inconsistent with physicalism as it is commonly understood.³

² In what follows, it will often be expositoryly convenient to refer to the *concept of F* alongside the *meaning of the term ‘F’*. For present purposes, these phrases can be viewed as equivalent. In particular, indeterminacy in the *meaning of ‘F’* will imply indeterminacy in the *concept of F*, and vice versa.

³ See the end of section 4 for citations of illusionists expressing this view.

A posteriori physicalists, on the other hand, deny that the concept of consciousness requires any features that would render the existence of consciousness inconsistent with physicalism. These physicalists acknowledge that there is a conceptual gap of some sort between the physical and the mental, but deny that this conceptual gap entails any metaphysical consequences. In this connection, they typically invoke ‘phenomenal concepts’ to account for the conceptual gap. They say that these are ways of thinking about conscious states which are directly recognitional and so have no a priori conceptual ties to any concepts of physical or functional states. Their directly recognitional nature rules out any possibility of inferring knowledge of conscious states from any physical or functional account of brain processes, but this does not mean that they do not in fact refer to physical or functional states. A posteriori physicalists thus view identifications of conscious states with brain states as akin to other a posteriori scientific identifications of everyday types with physical types, like the identification of water with H₂O or lightning with electric discharge.

Given this initial comparison of illusionism and a posteriori physicalism, it might seem surprising that the two camps should disagree so strikingly. Both agree on the underlying metaphysics—physicalism, understood in the traditional way. Moreover, illusionists generally agree with a posteriori physicalists on the existence of phenomenal concepts⁴, in that they too think that we have distinctive ways of thinking about conscious states that have no a priori ties to any concepts of physical or functional states. The difference arises only on the relatively specific issue of what our phenomenal concepts require of conscious states. The illusionists hold, while the physicalists deny, that requirements implying non-physicality are built into these phenomenal concepts.

⁴ Frankish 2016, pgs. 18-20 explores several routes an illusionist can take with regard to phenomenal concepts, for instance that phenomenal concepts might be “compounded from more primitive [concepts] that do refer, or that they have counterfactual causal connections to uninstantiated phenomenal properties”, or that they are hybrid concepts with a direct referential component which refers to a complex physical state and a theory-laden component which fails to refer. Kammerer 2021 develops an account under which “the content of phenomenal concepts, as determined by our theory of mind module, makes it impossible for us to think that it appears to us that we have an experience even though we don’t have the experience” (pg. 895). He further says that our subpersonal theory of mind module consistently causes us to judge that phenomenal properties are epistemically special, and this explains resistance to accepting the plausibility of illusionism.

2 Elimination or Reduction?

Illusionists say we should *eliminate* consciousness. They view our situation as akin to historical episodes which excluded certain entities from scientific discourse. Consider the nineteenth-century fate of *caloric*. As the history is normally viewed, when scientists in the nineteenth century found out no fluid flows from hot to cold bodies in heat conduction, they thereby found out that there is no caloric. They did not react to the discovery that heat is a kinetic phenomenon by saying, 'Oh, the stuff we've been calling "caloric" isn't a fluid after all, and we should now accept that "caloric" refers instead to the kind of kinetic energy involved in heat.' Instead they behaved as if the commitment to a fluid was an ineliminable element of the concept of caloric. According to illusionists, assumptions implying non-physicality are similarly built into the concept of consciousness, and so our recognition of physicalism implies there is no consciousness.

A posteriori physicalists, on the other hand, want to *reduce* consciousness. Their historical models are cases where new discoveries were taken to reveal more about existing entities, not to eliminate some category. Consider the property of being *a star*. Pre-modern astronomers thought that stars were fixed in celestial spheres which carried them daily around the earth. However, when modern astronomers came to recognize that stars lie in relatively fixed positions in space at many different distances from the rotating earth, they did not thereby conclude that there are no stars. In this case the association with celestial spheres was not regarded as built into the very idea of a star. Instead it was simply discarded as an incorrect assumption that did not fit the facts. According to a posteriori physicalists, this is what we should do with any assumption implying that consciousness is non-physical. We should not regard any requirement implying non-physicality as built into the concept of consciousness, and so open the way to the identification of conscious states with physical or functional states.

Given this comparison, one might think that it would be straightforward to adjudicate between illusionism and a posteriori physicalism. Is the concept of consciousness like the concept of caloric or like the concept of a star? That is, are the contested assumptions implying non-physicality built into the concept of consciousness, as the assumption of fluidity was arguably built into the concept of caloric, or is it a non-conceptual factual add-on, as the assumption of celestial spheres was arguably independent of the concept of a star?

However, this question assumes there is a fact of the matter about the concept of consciousness. Our central contention is that this assumption is unwarranted. Indeed, we take it that the appeal to historical examples itself rests on an assumption of conceptual determinacy that is unjustified. There is no reason to suppose that the requirement of fluidity was definitely built into the nineteenth-century concept of caloric, nor that the association with celestial spheres was definitely extraneous to the pre-modern concept of star. The history is equally consistent with its being indeterminate whether these assumptions were conceptually required, and the subsequent history being due to a free choice about how to resolve this indeterminacy. Nineteenth-century physicists could instead have deemed the requirement of fluidity to be extraneous, and come to view caloric as realized by internal kinetic energy, while early modern astronomers could equally have taken the view that stars are conceptually tied to celestial spheres, and so deemed radiant sources beyond the solar system not to be stars.

We take this to be a standard phenomenon in the development of human thought. At one time it is assumed that members of category A have property B. Then it is discovered that the items standardly categorized as A are not B. At this point the question is—reduce or eliminate? Should we conclude that the property A has turned out to *reduce* to a property that has turned out not to involve B after all? Or should we *eliminate* the property A on the grounds that instances of A are not possessed by B?

Our examples so far all fit this pattern. Should we reduce caloric to kinetic energy, or eliminate it because heat doesn't involve a fluid? Should we reduce stars to fixed distant radiation sources, or eliminate them because there are no celestial spheres? Should we reduce consciousness to physical processes, or eliminate it because nothing is non-physical?

3 Conceptual Indeterminacy

Our contention is that such choices are standardly undetermined. Nothing in prior usage determines which way to go. Prior to the discovery that candidate As are not B, nobody had any occasion to decide whether or not B was built into the concept of A. So, when the question is forced, neither option is ruled out.

Of course, if all concepts were nothing but directly referential, with no information of any sort built in, then choice between reduction and elimination would never arise. Reduction would

always be the correct option. After all, if the assumption that some B is required were always external to the concept of As, there would be no question of eliminating A because new information showed its members did not possess some property B.

However, it is doubtful any concepts ever work in a fully referential way. This is not to deny that such factors as causal origins, ostension and reference attraction can play some role in fixing concept reference. But these processes cannot determine a reference on their own. Apart from anything else, the ‘qua problem’ shows that subjects must have some idea of what *kind* of entity they are being connected to or pointed at, if they are to be credited with the ability to think about it. It’s no good just showing me a shoe, for example, and saying ‘*that*’. Such unadorned ostentation will leave me in the dark as to whether what’s at issue is the particular object before me, or a specific type of shoe, or shoes as such, or the leather material, or a particular kind of leather, . . .⁵

So at least some articulated assumption about the nature of the referent must be built into a concept if it is to secure a referent. And this is confirmed by the way that reduction is not always the automatic response to new discoveries about the candidate members of some category. When our intellectual ancestors decided that there is no caloric, or phlogiston, aether, witches, humours, and so on, they were not denying that the concepts involved had causal origins in real ostensible phenomena, but rather asserting that no part of those phenomena satisfied requirements that were built into the relevant concepts—such as: fluidity, absorbed in combustion, at absolute rest, supernatural powers, bodily fluids, and so on.

This now raises the question of exactly *which* assumptions constitute conceptual requirements on the referents of concepts. The central contention of this paper is that for many theoretically interesting concepts this will be an indeterminate matter. There is a principled reason for this—namely, that the effective use of concepts can happily tolerate indeterminacy. As a general rule, discourse can proceed objectively and efficiently even while leaving it open exactly what requirements are built into concepts. It is only in certain special circumstances that such indeterminacies become troublesome and force a choice between reduction and elimination. And then, given the prior indeterminacy, there will be no right or wrong about which way to go, but just the pragmatic question of how best to refine our concepts when some resolution is demanded.

⁵ Papineau 1979 ch 5 sect 7, Devitt and Sterelny 1987 ch 5 sect 3.

To understand why conceptual indeterminacy should be rife, note that the primary demand on concepts is that they should have unique referents in the actual world. As long as this requirement is satisfied, claims made using those concepts will have definite truth values. Now, there are two dangers in the way of securing such unique referents. First, the concept mustn't be so uninformative that it fails to fix on a definite referent. Second, it mustn't include so much information that it rules out all candidate referents.

We saw above how any concept that appeals only to ostension or causal origins will fall at the first hurdle. We need at least *some* information to tell us whether we are referring, for example, to a particular object, or a type of shoe, or shoes as such, . . . On the other hand, if we build *all* our factual beliefs into a concept, then we will fall at the second hurdle. Since there is no real possibility of our ideas being correct in every detail, building in all our assumptions would stop our concepts ever referring.

Still, between these two dangers there is plenty of room for indecision. Once we have built in enough information to leave us with only one candidate referent, it won't hurt if we add in a few extra assumptions, provided that those extra assumptions are indeed true of that referent. Building extra informational detail into the concept won't stop it referring if the extra information is accurate. The danger of eliminating any referent at all is only threatened by building in information that the candidate referent fails to satisfy.

Because of this, intellectual traditions will standardly have no reason to precisely pin down the informational content built into their concepts. To take a simple example from contemporary science, is it part of the concept *atom* that atoms must have nuclei? Of course it is *true* that atoms have nuclei. But is this part of the concept? Let us agree that science needs to assume that atoms are small material particles, one sort for each element, that combine by chemical means, in order to fix the referent of the concept *atom* to start with. But it can then happily leave it open whether or not having nuclei is also built into the concept of an atom. Given that it is effectively certain that atoms do have nuclei, reference to the same particles will be secured whether or not this extra assumption is part of the concept *atom*.

We might say of concepts like these that they are determinate in *extension* even though indeterminate in *intension*. The indecision about how much information is built into the concept

does mean that it is left open what they would refer to in certain other possible worlds taken as actual. If we were actually in a world where the small material particles distinctive of each element had no nuclei, it would be indeterminate whether our concept *atom* referred to those particles or not. But this does not mean that it is indeterminate what that concept refers to in the real world. Since the small material particles distinctive of each element do have nuclei in the actual world, our concept *atom* will refer determinately to those particles whichever way it is constituted.

So we can see why our intellectual predecessors might have seen no need to determine whether or not the requirement of fluidity was built into the concept of *caloric*, or that of celestial sphere embedding was built into the concept of *star*. To the extent that they were confident that the referent of their concept satisfied this assumption, they would have been happy to leave the issue open. As they saw it, the same entity would be picked out either way. Fussing about the precise analysis of their concepts would have seemed to them as silly as contemporary scientists fussing about whether nuclei are conceptually demanded of atoms. The same goes for the other examples we have already mentioned—was *absorbed in combustion* built into the concept of *phlogiston*? Was *supernatural* built into *witch*? *Absolute rest* into *aether*? *Bodily fluids* into *humours*? And further examples suggest themselves. Was *Euclidian* built into *straight line*? *Velocity-invariant* into *mass*? And so on.⁶

In all these cases, we maintain, the original users of these terms would have seen no need to answer these questions, since from their perspective the same entity would be picked out either way. It is only in retrospect that we can see they were wrong to be so insouciant. We now know that the real world is not as they supposed, but in fact one where the concept at issue is actually indeterminate in reference. Despite the confidence of previous thinkers, it turned out that the quantity involved in heat conduction is not a fluid; it turned out that those bright spots in the night sky are not embedded in celestial spheres. And then of course a choice was inevitable. The extension of the concept, and not just the intension, had now proved indeterminate, and so left the truth value of some statements open. It allowed some people to say ‘caloric is kinetic energy’, where others said ‘there is no caloric’. Again, some could say ‘stars are distant radiation sources at varying distances’, where others said ‘there are no stars’.

⁶ Quine held that *all* our assumptions are partly conventional and partly factual (1954 132). Our current claim is weaker and more plausible: for *some* assumptions, it is indeterminate whether or not they are required by the concepts they involve.

In situations like these, communication breaks down, and we need to agree to go one way or the other. But there is no right or wrong here. People might argue, but they are arguing about nothing. Nothing substantial is at issue. They agree about the nature of the world. They are just squabbling about how to refine a concept that their intellectual predecessors were wrongly confident could be happily left indeterminate.

So it is, we say, with consciousness. One might take the concept *conscious* to initially secure reference with the help of ostensive examples plus some such ideas as being private and introspectively accessible. But is *non-physical* also part of the concept? Until relatively recently it was generally assumed with confidence that the states picked out by the concept are all extra to the physical realm. So nobody would have seen any need to decide whether *non-physical* was part of the concept *conscious* or not. Since all the states in question were certainly non-physical, they thought, it wouldn't make any difference to the reference of the concept whether or not a requirement of non-physicality was imposed.

But now an appreciable body of opinion has come to the view that nothing in the spatiotemporal world is non-physical. And this then puts pressure on the concept *conscious*. Some say 'conscious states are physical brain states', where others say 'there is no consciousness'. As before, communication breaks down, and we need to agree to go one way or the other. But nothing substantial is at issue. We agree about the nature of the world and are only disputing how best to refine a previously indeterminate concept.⁷

4 Conceptual Indeterminacy in Consciousness

So it is, we say, with consciousness. One might take the concept *conscious* initially to have secured reference with the help of ostensive examples plus some such informational requirements as being internal and introspectively accessible. But were these requirements so strong as to render consciousness incompatible with physical realization? Until relatively recently it was generally assumed with confidence that the states picked out by the concept would all be extra to the physical realm.⁸ So nobody would have seen any problem about requirements that were incompatible with physicalism, and correspondingly nobody would have felt no need to decide

⁷ This section draws on the analysis in Papineau 1996.

⁸ For the prevalence of dualist views until the latter part of the twentieth century, see Papineau 2001.

whether they were part of the concept *conscious* or not. Since all the states in question were certainly non-physical, they thought, it wouldn't make any difference to the reference of the concept whether or not some requirement implying non-physicality was imposed.

But now an appreciable body of opinion has come to the view that nothing in the spatiotemporal world is non-physical. And this then puts pressure on the concept *conscious*. We saw above how the conclusion that there are no celestial spheres forced a decision between a weak and strong refinement of the concept *star*. On the weak refinement, the conclusion implied 'stars are distant heavenly bodies, not fixed in celestial spheres', while on the strong refinement it implied 'there are no stars'. In just the same way, the conclusion that all concrete things are physical forces a decision between a weak and strong refinement of *conscious*. On the weak refinement, the conclusion implies 'conscious states are brain states not extra to the physical realm', while on the strong refinement it implies 'there is no consciousness'.

And so, just as in the scientific cases, a decision is needed. Communication is in danger of breaking down. We need to go one way or the other. But nothing substantial is at issue. We agree about the nature of the world and are only disputing how best to refine a previously indeterminate concept.

Our analysis hinges on the assumption that illusionists are committed to a strong reading of the concept *conscious*. But so far we have not given details of this. In part this is because illusionists themselves are mostly content to leave their reading relatively open, claiming only that it suffices to rule out any physical realization.

Thus Keith Frankish 2016 says that phenomenal properties can be:

“. . . characterized as simple, ineffable, intrinsic, private, and immediately apprehended, and many theorists argue that they are distinct from all physical properties, inaccessible to third-person science, and inexplicable in physical terms.” p 17

And Georges Rey 2016, without further elaboration, attributes to most people:

“. . . a notion of ‘s(trong)-consciousness’, a notion that seems to me to give rise to Levine’s ‘explanatory gap’: it seems virtually impossible to imagine what further physical condition a w-conscious desktop should have to satisfy to become s-conscious.”

Again François Kammerer, in his 2018 paper ‘Can You Believe It?’, cites the just-quoted Frankish paper and says that phenomenal properties can be:

“. . . understood as *ineffable and irreducible qualities*; conscious experiences are said to have an *intrinsically subjective mode of existence*, and to be *directly and essentially apprehended* by the subject who has them. For all these reasons, it is very difficult to understand how something purely physical could be conscious: How can a purely physical brain, for example, give rise to such a peculiar thing as a *conscious experience*? p 46

This open-endedness on the part of illusionists complicates the dialectical situation somewhat. Note that the illusionist argument can be factored into two parts: (a) consciousness conceptually requires certain Φ s (ineffable, intrinsic, private, . . .), and (b) these Φ s cannot be physically realized—so in a physical world there is no consciousness. We have been taking the (b) part of this argument to be uncontentious and focusing instead on (a). However, it would always be open to an a posteriori physicalist to grant the (a) part with respect to some Φ , and then counter that this Φ can in fact be physically realized. Still, this is not the line normally taken by a posteriori physicalists, as opposed to contesting whether illusionists are right to build their Φ s into the concept of consciousness. This makes sense given the open-endedness of the illusionist analysis of the concept. Even if an a posteriori physicalist could argue that some specific Φ is physically realizable—immediate apprehensibility, say—there would remain other Φ s—ineffability, irreducibility, . . .—for which this line would not be viable. In line with this, we shall continue to focus on the (a) part of the illusionist argument, and to query whether they are right in their strong claims about the conceptual requirements for consciousness.

5 Conceptual Indeterminacy in Philosophy

So far, we have argued that many theoretically interesting terms are indeterminate in meaning. Confidently held background assumptions will often imply that the users of a term have no reason to decide exactly which descriptive specifications are built into its meaning. So when these background assumptions are violated, and it turns out that the putative referent of the term does not possess a feature which users had so far been attributing to it, they are confronted with a decision: reduce or eliminate. Neither option is forced by facts in the world nor the meaning of the term.

We have applied this view to the term 'conscious', as well as to scientific terms like 'caloric', 'star' and others. One might wonder whether there are philosophical terms other than 'conscious' for which this treatment may be applicable. As a matter of fact, we think philosophy is riddled with such indeterminate terms. We suspect that indeterminacy of this sort may even be a defining feature of a large number of perennial philosophical disagreements, since this sort of indeterminacy entails that there is no decisive way to resolve disagreements. In this section, we will apply our analysis to the term 'free will'. In addition to being of independent interest, this example of further conceptual indeterminacy in philosophy should lend credence to our analysis of 'conscious'.

Consider the term 'free will'. Let us suppose that the original users of the term took it that free will required *uncoerced decisions*, in the sense of some paradigmatic sort of human decision-making events, and that they also took it that free will required *not governed by physical law*. But was the latter requirement built into the meaning of 'free will'? To the extent that the original users of the terms were confident that all uncoerced decisions lay beyond the domain of physical law, they would have had no reason to resolve this question. Exactly the same set of actual actions would come out as freely willed either way.

Modern science, however, argues strongly that nothing is beyond the realm of physical law, not even uncoerced decisions. So now we face a choice. If we assume the hegemony of physical law, does it follow that humans lack free will, or not?

We think there is no decisive answer to this question. Nothing in the original usage decided whether *not governed by physical law* was built into the concept of free will alongside *uncoerced decisions*. However, in the face of the possibility that everything is law-governed, a decision is needed. Either we regard *not governed by physical law* as built into the concept, and so regard the

ubiquity of physical law as eliminating free will. Or we regard it as nothing more than a false assumption about free will upheld by our intellectual ancestors, and so are able to reduce free will to uncoerced decisions occurring within a law-governed world.

Which way of understanding 'free will' is correct? Incompatibilists—who hold that free will requires transcending physical law—urge that the former definition of 'free will' is correct; whereas compatibilists—who take free will to be consistent with being governed by law—hew to the latter. In our view, neither party has a stronger claim. This strikes us as another instance in which philosophers are disagreeing over what a term means, even though nothing in prior usage dictates an answer. It is ultimately a matter of free linguistic decision whether or not we decide that 'free will' requires independence of physical law.

So in our view the debate between incompatibilists and compatibilists is not a substantial one. It does not hinge on any disagreement about the nature of the world, but just about how best to refine an indeterminate concept when the need for such refinement becomes apparent.

Of course there are substantial issues in the vicinity, issues which certainly involve more than mere terminological decision. In particular, there are questions about how our moral and legal practices should respond to the possibility that humans and their activities fall within the realm of physical law. Does the hegemony of physical law imply that nobody should be blamed and punished, or praised and rewarded, for their actions? This clearly isn't just a terminological question, but a substantial issue for moral and legal debate.

But note that this substantial question can be posed whichever way we refine 'free will', or indeed without using the term at all (as we did in the last paragraph). Thus, assuming still the hegemony of physical law, the incompatibilist could ask 'If we agree there is no free will, does that mean nobody should be blamed and punished?', while the compatibilist could similarly ask 'If we agree that all free actions are law-governed, does that mean nobody should be blamed and punished?'. Indeed, we can all ask (to keep things simple) in the style of the last paragraph, 'If we agree that all human choices are law-governed, does that mean nobody should be blamed and punished?'. To hold that disagreements about free will are empty does not require us to take the same view about all associated issues. On the contrary, it enables us to separate the substantial issues worth focusing on from empty debates about terminology.

We take the example of free will to carry a general moral. As we said, we think that conceptual determinacy is rife within philosophy. Many philosophically significant terms, including ‘person’, ‘belief’, ‘morality’, ‘rationality’, and others, leave it open exactly what they conceptually require of their referents. The result is that much philosophical debate can get bogged down in terminological disagreement and be distracted from the important issues. Awareness of this danger might do much to remedy it.

6 Objections

Our view, then, is that the contemporary debate between the illusionists and a posteriori physicalists is yet one more instance where an essentially terminological choice is being mistaken for a substantial dispute. Let us conclude by addressing some possible objections to this diagnosis.

1. Indeterminacy Assumes Physicalism. Should not the central claim in this paper be put as a conditional, such as ‘if orthodox physicalism is true, then it is indeterminate whether the term “conscious” refers or not’? After all, haven’t we been assuming the plausibility of an orthodox version of physicalism, and been urging that the relatively recent rise to prominence of this view is the reason why we are now confronted with the decision to reduce or eliminate consciousness? On this interpretation of our central claim, someone who was *not* a traditional physicalist—like a dualist or Russellian monist—might say that our consciousness concept is determinate, and that it refers to properties which are outside the scope of orthodox physicalism.

However, we think that this interpretation of our central claim is incorrect, and that the term ‘conscious’ is indeterminate regardless of whether traditional physicalism is true. Our contention is that it is indeterminate whether or not requirements incompatible with physicality are built into the meaning required of ‘conscious’ states. If orthodox physicalism were false, then it is true that there would be no pressing need to decide this question, since the paradigm candidates for conscious states would qualify as conscious either way, given that they would be non-physical. On the assumption of non-physicalism, then, the question of whether to reduce or eliminate ‘consciousness’ would not arise, and the indeterminacy could be left unresolved. But for all that the term ‘conscious’ would still be indeterminate in definition, since it would still be an open question of whether to reduce or eliminate consciousness *if* physicalism turns out to be true.

2. *Substantial Issues that are not Indeterminate.* Are there not substantial issues about consciousness that cannot simply be dismissed as due to the supposed indeterminacy of the term 'conscious'? Yes indeed. Just as with free will, we accept that there are substantial questions in play which transcend any indeterminacy in our concepts. In particular, there are questions about the soundness of the knowledge argument and the conceivability argument against orthodox physicalism.

Still, we think these questions can be posed without invoking any putatively indeterminate concept, in the way we might avoid the term 'free will' in asking whether or not we ought to be blamed for our actions if they are completely subject to physical law. Thus we might wonder whether Mary learns about some new feature of reality when she leaves her black-and-white room and sees red for the first time, or whether our perfect physical duplicates would necessarily be full duplicates simpliciter. Notice that, so articulated, these puzzles do not mention 'consciousness' at all.

3. *Physicalism is Implicit Illusionism.* Are not a posteriori physicalists open to the charge of being closet illusionists? Their opponents will point out that the 'phenomenological concept strategy' for defending physicalism is typically accompanied by arguments holding that the special structure of phenomenal concepts leads people to think consciousness is non-physical. When combined with an acceptance of physicalism, this looks very close to accepting illusionism. After all, if our phenomenal concepts inevitably lead us to judge that consciousness is non-physical, is this not good grounds for thinking that it is built into our 'consciousness' concept that consciousness is non-physical?

We agree that people are strongly inclined to judge that consciousness is non-physical—perhaps because of the special structure of phenomenal concepts—but we do not agree that this induced assumption is sufficient to show that non-physicality is conceptually required of 'consciousness'. Why shouldn't 'consciousness' be indeterminate even if we are strongly inclined to judge that consciousness is non-physical? Consider 'star' once more. We take it that it was very intuitive for humans to judge that stars—especially the sun—orbit the earth in fixed celestial spheres. Yet, despite being strongly inclined to judge that stars have this property, the concept was indeterminate with regard to whether it conceptually required fixed celestial spheres. Similarly, our internal cognitive machinery might strongly incline us to hold that the concept of

consciousness requires non-physicality, yet it be indeterminate whether this implication is built into the concept itself.

4. *The Phenomenological Fallacy*. A similar point applies to the apparent involvement of worldly properties like yellowness and squareness in certain conscious states. As U. T. Place originally observed (1956), it is very natural to think that sensory experiences literally incorporate the worldly properties they respond to, and thence to conclude that experiences must be non-physical, given that physical brain processes involved in seeing yellowness and squareness are not themselves yellow or square. (Place called this “the phenomenological fallacy”.)

Illusionists often appeal to this line of thought in defence of their position. They argue that the intuitive involvement of worldly properties in conscious experiences show that our concept of consciousness rules out physical realization.⁹ But our response to this objection is the same as to the last one. It might be natural to think that conscious experiences incorporate worldly properties, and the incompatibility with physicalism might immediately follow. But this does not establish that ‘conscious’ semantically imposes requirements incompatible with physicalism, any more than the naturalness of thinking that stars are fixed in celestial spheres means that this assumption is semantically required by the term ‘star’.

5. *Ensouledness Trumps Physicality*. What if it were to turn out that the human population were extremely heterogeneous when it comes to mentality? To illustrate with an extreme case, suppose that some humans had non-physical souls which conferred mentality on them, while other people were purely physical beings. One might feel that in this scenario only the people with souls would have conscious mental lives, and the purely physical beings would not. And this would then argue that ‘conscious’ determinately requires non-physicality, in order to explain why the purely physical beings would be denied consciousness.

However, we say that this line of argument begs the question. To the extent that the purely physical beings manifest the signs of paradigm conscious states, like itches and pains and sensory experiences, along with their ensouled counterparts, then there seems no immediate reason to deny them consciousness, unless we are already assuming some illusionist understanding of

⁹ For instance, Dennett 2016, pg. 70 suggests that the “seeds of illusionism can already be discerned in U.T. Place’s pioneering article, ‘Is Consciousness a Brain Process?’ (1956). Place was so bold as to identify the denial of illusionism as a fallacy, the phenomenological fallacy.”

‘conscious’. Of course, on our assumption that it is ultimately indeterminate what the meaning of ‘conscious’ is, it would be open to us to refine the term so that only the ensouled members of the population qualify as conscious. But we take it that nothing in the existing meaning of the concept forces this resolution.

6. *Animal Consciousness*. A more general worry is that we have made the term ‘consciousness’ too indeterminate. For instance, doesn’t it follow from our treatment that it is indeterminate whether cats, say, are conscious? We say not. This sort of indeterminacy does not follow from the particular claims we have made here. We say that ‘consciousness’ is indeterminate with regard to whether it imposes requirements incompatible with physicalism or not. It does not follow that it does not imply a definite answer on whether cats or other organisms are conscious. Thus the concept may specify clearly that the kind of action control and environmental sensitivity displayed by cats would be sufficient for consciousness—as long, that is, as cats satisfied any other requirements on consciousness assumed by illusionism. Then the implication would be that cats are definitely conscious—supervaluationally conscious, as it were—whether or not the illusionist’s requirements are conceptually built into the concept. Whichever way the concept is refined, cats would come out conscious all right. (Of course, it is also consistent with our position that the concept *conscious* does not lay down clear criteria which will so supervaluationally decide on consciousness for all organisms. Perhaps the concept does leave it open whether shrimps are conscious, or indeed whether cats are conscious. But, if so, these indeterminacies would not follow from the concept leaving it open whether or not strong illusionist requirements are built into consciousness, but would rather be additional dimensions of indeterminacy.)

7. *Pragmatic Issues*. We say there is nothing to decide between ‘consciousness is conceptually incompatible with physicalism and so there is no consciousness in a purely physical world’ and ‘consciousness exists even in a purely physical world, because it does not conceptually demand non-physicality’. In our view these are simply two equally admissible ways of resolving a terminological indeterminacy.

But are such eliminative and reductive options really always both equally admissible, even on the assumption that the term at issue is semantically indeterminate? Consider ‘witch’. Surely it is implausible that, if the witch-persecutors of Salem in 1693 had become convinced that the women on trial lacked supernatural powers, they would have continued applying the term ‘witch’ to them? Even if the term ‘witch’ had been technically indeterminate, surely there would have

been extremely strong pragmatic pressures to conclude the women on trial were not 'witches', once it was agreed they had no magical powers.

Perhaps some illusionists will want to defend their position along these lines. They might try to persuade us that there are strong pragmatic reasons for preferring their option. Given all the surrounding attitudes and assumptions, a better overall result will be achieved by eliminating consciousness rather than reducing it.

Maybe so. But this is a different argument. Given that we will need to choose between reduction and elimination, if we are to continue talking of 'consciousness', perhaps there is indeed a pragmatic case for preferring eliminativism. As it happens, we doubt it. It strikes us as rather more sensible to try to come to terms with the idea that conscious states are physical than to opt for saying that there is no consciousness. But, in any case, this is now an issue that takes us beyond the focus of this paper. Frankly, we don't know how the pragmatic issues will play out, and further, we have made no claim to know this.

It suffices for the argument in this paper that there be no non-pragmatic *fact of meaning* which forces either reduction or elimination. If there are pragmatic reasons to prefer one route over the other, so be it. But a pragmatic reason for preferring one way of refining an indeterminate term is not the same as an argument vindicating one side of the debate between illusionism and a posteriori physicalism. Our analysis implies the search for such an argument is doomed to failure and that the parties to the debate will do well to abandon it.

References

Braddon-Mitchell, D. 2003 'Qualia and analytical conditionals'. *Journal of Philosophy* 100 (3): 111-135.

Chalmers, D. 1995 *The Conscious Mind* Oxford: Oxford University Press

Churchland, P. 1981 'Eliminative Materialism and the Propositional Attitudes' *Journal of Philosophy* 78: 67-90

Davidson, D. 1973 'Radical Interpretation' *Dialectica* 27: 313-28

- Devitt, M. and Sterelny, K. 1987 *Language and Reality* Cambridge Mass: MIT Press
- Dennet, D 1971 'Intentional Systems' *Journal of Philosophy* 68: 87-106
- Dennett, D. 2016 'Illusionism as the Obvious Default Theory of Consciousness' *Journal of Consciousness Studies* 23 (11-12): 65-72
- Fodor, J. and Pylyshyn, Z. 1988 'Connectionism and Cognitive Architecture' *Cognition* 28: 3-71
- Frankish, K. 2016 'Illusionism as a Theory of Consciousness' *Journal of Consciousness Studies* 23 (11-12): 11-39
- Johnston, M. 1992 'How to speak of the colors'. *Philosophical Studies* 68 (3): 221-263.
- Kammerer, F. 2018 'Can you believe it? Illusionism and the illusion meta-problem' *Philosophical Psychology* 31 (1): 44-67.
- Kammerer, F. 2021 'The Illusion of Conscious Experience' *Synthese* 198: 845-866.
- Lewis, D. 1995 'Should a materialist believe in qualia?' *Australasian Journal of Philosophy* 73 (1): 140-44.
- Niikawa, T 2021 'Illusionism and Definitions of Phenomenal Consciousness' *Philosophical Studies* 178:1-21
- Papineau, D. 1979 *Theory and Meaning* Oxford: Clarendon Press
- Papineau, D. 1996 'Theory-Dependent Terms' *Philosophy of Science* 63: 1-20
- Place, U. 1956 'Is Consciousness a Brain Process?' *British Journal of Psychology* 47: 44-50
- Quine, W. 1954: 'Carnap and Logical Truth' reprinted in his *The Ways of Paradox and Other Essays* New York: Random House 1966 100-125

Ramsey, W., Stich, S., and Garon, J. 1990 'Connectionism, Eliminativism and the Future of Folk Psychology' in Tomberlin, J. ed *Philosophical Perspectives* 4: 499-533

Rey, G. 2017 'Taking Consciousness Seriously-- as an Illusion' reprinted in Frankish, K. *Illusionism: As a Theory of Consciousness* Imprint Academic