
*Causal Essentialism versus the Zombie Worlds*¹

BRIAN JONATHAN GARRETT
McMaster University
Hamilton, ON L8S 4L8
CANADA

David Chalmers² claims that the logical possibility of ‘zombie worlds’ — worlds physically indiscernible from the actual world, but that lack consciousness — reveal that consciousness is a distinct fact, or property, in addition to the physical facts or properties.

The ‘existence’ or possibility of Zombie worlds violates the physicalist demand that consciousness logically supervene upon the physical. On the assumption that the *logical* supervenience of consciousness upon the physical is, indeed, a necessary entailment of physicalism, the existence of zombie worlds implies the falsity of physicalism. How do we determine the logical possibility of zombie worlds? By conceptual analysis of the concepts involved, keeping empirical facts in mind.³ Keeping our methods in mind, we can therefore articulate the argument like this:

-
- 1 Thanks to Torin Alter, Andrew Bailey, Phil Kremer, Antonia LoLordo, William Seager, Mark Walker and the audiences of the APA Pacific Division and the CPA Halifax, for discussion on an earlier version of the paper. Special thanks to the anonymous reviewers for their sharp and useful comments.
 - 2 David J. Chalmers *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press 1996)
 - 3 See Andrew Melnyk ‘Physicalism Unfalsified: Chalmer’s Inconclusive Conceivability Argument’ in *Physicalism and Its Discontents*, Carl Gillett and Barry Loewer, eds. Cambridge: Cambridge University Press 2001.

- (i) I can conceive of two physically indiscernible possible worlds, one of which (the actual world) includes consciousness, but the other does not.
- (ii) If I can conceive of these worlds, then (defeasibly) these worlds are logically possible.⁴
- (iii) If it is logically possible that there are two worlds physically indiscernible, one of which lacks consciousness, then consciousness is not a physical fact/property.

Therefore:

- (iv) Consciousness is not a physical fact/property [non-reductivism].

Although I like the conclusion, I shall argue skeptically that alternate metaphysical assumptions undermine Chalmers's argument. I shall argue that if we accept causal essentialism (CE) regarding physical property identity, then the worlds Chalmers conceives are not, in fact, logically possible. The intuition that these worlds *are* possible is a result of not keeping all our appropriate metaphysical commitments firmly in mind. Put in a less partisan way, Chalmers's zombie worlds are only genuine possibilities for Humeans regarding laws of nature and property identity. I shall argue that if we reject a Humean approach to laws and properties, accepting modal or causal essentialism, then we cannot accept the genuine possibility of zombie worlds.⁵

4 Premise (ii) If I can conceive of these worlds, then (defeasibly) these worlds are logically possible, is in my mind preferable to this counterpart (ii*) If the zombie world is ideally conceivable then the zombie world is possible. Talk of 'ideal' conception, like ideal epistemic situations generally, tend to be difficult to explicate without circularity, and they are nothing like idealizations in scientific theory.

5 I say 'modal or causal' to note that I do not wish to commit myself to the existence of fundamental causal dependencies. There may be law, or counterfactual dependence, without causation relating the most fundamental entities. Those who assume that the basic relations will be causal or that the fundamental laws will be causal laws, hold a particularly strong form of physicalism. But causation is an old and contested metaphysical subject — and the phenomena of causation might not be in need of saving. There may be no need to understand the fundamental laws as causal — and if the failures to analyze causation successfully lead to pessimism regarding its existence, we may have some reason to doubt that they are causal. However, my arguments should work for those who believe that causation is found between the fundamental physical properties.

Properly conceiving the physical is as important as properly conceiving the phenomenal. Here the emphasis is on the physical. Current and popular views regarding the nature of physical concepts and properties as 'functionalizable,' accepted by a wide variety of philosophers, can be explained if causal essentialism is true. Causal essentialism, therefore, offers a plausible conception of the physical, yet causal essentialism causes trouble for nonreductivists who appeal to the logical possibility of zombies⁶.

In section one I introduce causal essentialism and its defenders, along with the zombie argument. I note, very briefly, some of the advantages causal essentialism has over Humean alternatives. But the reader looking for a thorough defense of causal essentialism will be disappointed. Rather, I mention some of the advantages of the theory and I offer some new considerations for favoring the theory, but I leave a full defense to others.⁷ In section two I discuss how causal essentialism can offer us a natural account for why physical property concepts seem to be causal-role or functional concepts, as Kim and Chalmers claim.⁸ I argue that Chalmers's comments regarding complete explanations of physical properties are well accounted for by causal essentialism. Thus I am urging that causal essentialism regarding the physical is a natural metaphysical view for many, dualists and materialists alike. In section three I return to the incompatibility of zombie worlds with causal essentialism and in section four I turn to a number of objections and clarifications and finally, in section five, I briefly discuss the relationship between essentialism and dualism. Although causal essentialism deprives the dualist of the zombie possibilities, causal essentialism itself is consistent with at least one possible, although unlikely, nonreductivist view about phenomenal properties.

Whether we should accept causal essentialism depends on that view's success at explaining other metaphysical puzzles and whether we can in all good intellectual conscience accept its consequences. CE is part and parcel of a general metaphysical picture, one of many competing speculative systems, but I shall advocate it is a live metaphysical option

6 However, Jackson's knowledge argument and arguments from multiple realizability do not rely on the logical possibility of zombie worlds, so they will remain unaffected by the argument of this paper.

7 See Brian Ellis, *The Philosophy of Nature: A Guide to the New Essentialism* (Montreal: McGill Queens University Press 2002).

8 Jaegwon Kim, *Physicalism, or Something Near Enough* (Princeton, NJ: Princeton University Press 2005); *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, (Cambridge, MA: Bradford, MIT Press 1998).

and one that best suits a plausible philosophy of science and provides us with a plausible conception of the physical. Naturally, a full defense of causal essentialism isn't possible here, (and it has been done well by others, notably by Brian Ellis), but showing that a major contender in metaphysical speculation is incompatible with Chalmers's conclusions is an important result. Those seeking to defend the ontological distinctness of phenomenal properties from physical properties must find another argument, one that is not so partisan, i.e. one that does not force us to decide between what are otherwise compatible options, a plausible non-Humean metaphysics of the physical versus an argument for irreducibility.

I Modal Essentialism and Laws of Nature

Consider two worlds and the following claims. For simplicity's sake there are only two objects in these worlds, a and b, and let us consider 'F' and 'G' to be the only physical properties of this world.⁹

w1	Fa and Gb For all x [F(x) -> there is a y, G(y)]
w2 (actual)	Fa, Ca and Gb For all x [F(x) -> there is a y, G(y)] For all x [F(x) -> C(x)]

Assume that w2 is the actual world and w1 the zombie world. First note that in w1 and w2 we are conceiving the very same two objects, or their counterparts, a and b. Second, both w1 and w2 are indiscernible regarding the *physical* law: For all x [F(x) -> there is a y, G(y)]. Third, the worlds differ regarding their properties; w2, the actual world, has a property C instantiated — some phenomenal property or qualia — which w1 lacks. But more needs to be said here. Chalmers conceives of the supervenience of consciousness upon the physical to hold across only those worlds that share the same laws of nature as the actual world. The zombie world is a logically possible world that lacks some of the laws of the

9 As stated, my worlds have only two physical properties and two objects and that may seem more confusing than simplifying, after-all, actual conscious persons are made of many parts and appear to have many physical properties. On the other hand, a 'unified field theory' might have ontological commitment to only one physical field and the only individuals of metaphysical worth might be only space-time points. In short, feel free to drop the 'simplifying' assumption.

actual world. The psycho-physical law: 'For all x [$F(x) \rightarrow C(x)$] is absent from the zombie world. Fourthly and finally, let us assume that there are no other laws underwriting these, such that these are fundamental or basic laws.

We thus appear to have a scenario required to make sense of Chalmers's thought experiment. w_1 and w_2 are physically indiscernible worlds, sharing as they do the very same physical objects, a and b , and sharing all the same physical laws: For all x [$F(x) \rightarrow$ there is a y , $G(y)$]. Consider yourself ' a ' in w_2 . We can now accurately conceive another world, physically indiscernible from the actual world, in which ' Fa ' is true, but ' Ca ' is not true; that's w_1 , the zombie world. So far we appear to have two logically and metaphysically possible worlds. But I am skeptical. The possibility of two such worlds depends on an assumption regarding the relation between *physical* property identity and the laws of nature. My concern is this: why think that these two worlds share the same *physical* properties? They may do so only on the assumption that a physical property-instance $P(a)$, can have the causal capacity (as described by the laws) to instantiate P^* in one world, and not have the capacity to instantiate P^* in a distinct, metaphysically possible world. This implies that the *identity* of a physical property-instance must *not* be essentially tied to its actual causal capacities (as described by the laws) to instantiate other properties¹⁰. Let me refer to this as 'Hume's thesis' since Hume is famous for his denial of necessary connections.

At first Hume's thesis seems to be a reasonable assumption, after-all the Humean intuition that all individual matters of fact are logically independent of one another appears reasonable. But since the postulation of a posteriori identities in Kripke's work, support for Hume's thesis has weakened. A modal claim based on conceptual intuition may be defeated by the plausibility of an a posteriori identity, requiring that our conceptual intuitions be backed up by some serious empirical investigation. We cannot confidently claim, based on a priori conceptual intuitions, that each fact is logically independent of every other. However, I don't intend to attempt a direct refutation of Hume's thesis, but rather to show that an alternative is viable and plausible.

So in contrast to Hume's thesis we may champion modal essentialism: that the modal properties, typically the causal capacities of physi-

10 Whether the fundamental laws are causal or not, is not something that I wish to presuppose as essential to physical properties. But this is an epistemological consideration due to our not yet knowing how best to articulate the basic laws. If they are indeed causal, as many metaphysicians assume, then my nomological essentialism will be a causal-nomological essentialism. However, I am skeptical that the fundamental laws count as causal.

cal properties are essential to those physical properties. The idea that the actual causal capacities of a property are essential to that property is not new.¹¹ In the 1950s physicist David Bohm wrote:

the qualitative causal relationship that water becomes ice when cooled and steam when heated is a basic part of the essential properties of the liquid, without which it could not be water. Similarly, the chemical law that hydrogen and oxygen combine to form water is a basic property of the gases hydrogen and oxygen, without which they could not be hydrogen and oxygen.¹²

Varieties of this view have been ably defended and developed by Sidney Shoemaker and independently by Peter Achinstein.¹³ As Shoemaker puts the thesis, 'all of the causal potentialities possessed by a property at any time in the actual world are essential to it and so belong to it at all times and all possible worlds' (Shoemaker, 222).

More recently, Brian Ellis has defended a metaphysical view he dubs 'scientific essentialism' that has, as its central metaphysical tenant, the claim that the laws of nature are not external to the properties those laws relate, but are rather logically supervenient upon the dispositional properties of the actual world. Ellis dubs his general view 'scientific essentialism.'

'The scientific essentialist holds (a) matter is not passive, but essentially active and reactive, (b) that the essential properties of things belonging to natural kinds include dispositional properties — that is, causal powers, capacities and propensities, (c) that the basic laws of nature are not descriptions of behavioral regularities, but of the ways in which things belonging to natural kinds must be disposed to act or interact, given their essential properties, (d) that the causal laws of nature are metaphysically necessary, because anything that has the essential dispositional properties of a given natural kind must be disposed to behave as these properties require, and (e) that elementary causal relations involve necessary connections between events — namely, between the triggers and displays of basic dispositional properties' (Ellis, 106).

11 There are several objections to the thesis based on the proliferation of necessities. Worlds must share all the same physical properties and laws or be entirely alien in their contents and physical possibility 'becomes' a species of logical possibility. But I think biting these bullets is to bite blanks, so no intellectual harm is done.

12 See David Bohm, *Causality and Chance in Modern Physics* (Philadelphia: University of Pennsylvania Press 1957).

13 See Sydney Shoemaker, *Identity, Cause and Mind: Philosophical Essays* (London: Cambridge University Press 1984); Peter Achinstein, 'The Identity of Properties,' *American Philosophical Quarterly* 11 (1974) 257-76.

Ellis argues that the advantage of his view regarding the essential relation between dispositional properties and laws of nature is that it offers answers to puzzles regarding the ontology of laws. Laws support counter-factual conditionals and this may be explained by the logical supervenience of laws upon the fundamental dispositional properties of the world. If the laws of nature logically supervene upon dispositional properties, then there is no world in which that physical property does not 'obey' the law. Thus the truth of a particular counterfactual that is supported by a law can be accounted for by the general metaphysical fact that the laws are metaphysically necessary.

Law statements also have a truth-maker under this view — they are true in virtue of their accurate description of the dispositional properties of the actual world. Compare the problem for a Humean regarding laws of nature — not all universal generalizations are laws, so the search for the ontological ground that distinguishes law-like from non law-like generalizations is pressing for the Humean. Scientific or causal essentialism here offers a nice unified solution to these problems.

II Causal Essentialism, Causal-role Concepts and the Limits of Science

Before turning to the alleged possibility of zombie worlds, it's worth seeing how causal essentialism accounts for physical properties and whether it accords with how physical properties are commonly conceived. Causal essentialism fits beautifully with functionalist or causal role accounts of physical natural kinds. Kim's account of reductively explaining a property invokes the idea of a functional definition in his first step.

'Step 1. [Functionalization of the Target property]
property M to be reduced is given a functional definition of the following form:

Having M=def. having some property or other P (in the reduction base domain) such that P performs causal task C.

Step 2. [identification of the realizers of M]
Find the properties (or mechanisms) in the reduction base that perform the causal task C.

Step 3. [Developing an explanatory theory]
Construct a theory that explains how the realizers of M perform task C.' 2005 p. 101

Now Kim thinks that functionalization is successful for physical properties and all special science properties with the exception of qualia. With regard to qualia he believes that functional definitions will not completely capture the phenomena of consciousness. David Chalmers writes

Even a physical notion such as heat can be construed functionally: roughly heat is the kind of thing that expands metals, is caused by fire, leads to a particular sort of sensation, and the like. Once we have an account of how these various causal relations are fulfilled, then we have an account of heat. Heat is a causal-role concept, characterized in terms of what it is typically causes, under appropriate conditions. Once empirical investigation shows how the relevant causal role is played, the phenomenon is explained. (45)

Two things are worth noting — the concept ‘heat’ is a causal role concept, and an ontological claim regarding what heat is. Causal essentialism may also hold that the concept of ‘physical property’ is a causal role concept.¹⁴ Empirical investigation reveals how the relevant role is played for each physical property and thus each physical property is explained. Basic physical properties will be those in which no further mechanisms can be found to fill the causal role. We may define a basic property as one that does not have any underlying causal mechanism. For example, if instances of phenomenal consciousness are not realized by some mechanism, then consciousness is not reductively explained. Vitalists are those who believe that no ‘lower-level’ property, such as the mechanisms of metabolism and reproduction, can be found that plays the role of life. The failure of vitalism is due to our empirical success in finding just such mechanisms in all examined examples of life.

Chalmers comes close to identifying physical properties as external role properties, leaving us with little to say regarding what intrinsic nature could be playing the role. While exploring some metaphysical possibilities for the mind-body problem he remarks on the nature of physical properties:

The strategy to which I am most drawn stems from the observation that physical theory only characterizes its basic entities relationally, in terms of their causal and other relations to other entities. Basic particles, for instance, are largely characterized in terms of their propensity to interact with other particles. Their mass and charge are specified, to be sure, but all that a specification of mass ultimately

14 This appears to be a very popular assumption among physicalists and dualists alike: popular for dualists who wish to contrast the essentially causal nature of the physical with the allegedly non-causal nature of the qualitative aspect of consciousness, and popular with physicalists who wish to find a ground or mechanism for causal relations between supervenient events or property instances.

comes to is a propensity to be accelerated in certain ways by forces, and so on... The picture of the physical world that this yields is that of a giant causal flux, but the picture tells us nothing about what all this causation relates. (153)

Chalmers then considers whether the intrinsic properties that play the roles specified could themselves be phenomenal or 'proto-phenomenal' properties. I am not concerned with this metaphysical picture, but it is important to note that *physical* properties accessible to human inquiry are causal-relational properties in this view. Indeed, Shoemaker in his defense of causal essentialism takes one of its advantages to be that it allows physical properties to be epistemically accessible to human agents, which most physicists appear to believe. However, Chalmers backs off, declaring that the intrinsic nature of physical properties is unknown to us: 'After all, we really have no idea about the intrinsic properties of the physical. Their nature is up for grabs, and phenomenal properties seem as likely as any other' (153).

Causal essentialists can agree that there may be some 'further' intrinsic nature to physical properties, over and above their causal capacities, but the causal essentialist will insist that the causal capacities of fundamental physical properties cannot be abstracted from the intrinsic character of these properties. The dispositional features of a physical property are as essential to that property as its alleged intrinsic or categorical nature.

Although it is orthogonal to my concerns, we might feel Chalmers is treading on thin ice — the more we assert that intrinsic nature of the physical is unknown to us, the more we lose a grip on how the argument for non-identity is supposed to proceed; after-all, if we know next to nothing about what the physical *really* is, then arguments regarding consciousness' distinctness from the physical will run afoul of the objection that the contrast drawn between the physical and consciousness takes non-essential, (yet epistemically available) aspects of the physical as its model, instead of the intrinsic and essential aspects of the physical as its model.¹⁵ Thus, if the intrinsic nature of the physical is distinct from its causal capacities to instantiate other properties, we allow these

15 I shall return to this point below. For example, if physical properties have causal capacities contingently (i.e. there are worlds in which physical properties occur, but don't share the same laws of nature as ours, hence they have either different causal capacities or no causal capacities), then they are not so different from those who claim that phenomenal properties, when causally efficacious, are so only contingently. The typical contrast drawn between phenomenal and physical properties is that (i) the phenomenal is non-relational but physical properties are not, and (ii) the phenomenal is qualitative, but physical properties are not, and (iii) phenomenal properties are epistemically immediate, but physical properties are not.

properties to be instantiated in some possible worlds with distinct capacities or even no causal capacities at all. If physical properties are not essentially causal, then they do not differ from consciousness so much. If physical properties do not have their causal capacities essentially, then they are on that point, much like consciousness is alleged to be, and that weakens the case for thinking them to be members of a distinct class of property.¹⁶ The worry might become this: if physical properties have intrinsic natures and it is logically possible that physical properties are epiphenomenal, then consciousness might be one of those epiphenomenal, yet physical properties. Kim and Chalmers agree that physical property instances are caused by physical property instances and are not over-determined; hence any allegedly nonphysical property, such as consciousness, is epiphenomenal. This 'exclusion argument' takes nonidentity with the physical as a premise and concludes that any such property is epiphenomenal. That argument isn't being rejected here, but the zombie argument intended to establish nonidentity is.¹⁷

Returning to the explanatory importance of essentialism, Chalmers is close to thinking of physical properties as having their causal capacities essentially. Discussing the explanation and ontology of heat, Chalmers writes:

...many causal-role concepts are partly characterized in terms of their effect on experience: for example, heat is naturally construed as the cause of heat sensations. Does this mean that we have to explain heat sensations before we can explain heat?...To be sure, no explanation of heat will be complete until we have an account of how that causal connection works, but the incomplete account is good enough for most purposes. (45)

-
- 16 I have no sympathy for the notion that consciousness is not essentially causal. The essential activity of soul, pneuma, life, thought, Fichtean absolute self, or whatever etc., is remarked upon by many authors. They might be wrong, but it is very hard to believe, given the history of these ideas, that the meaning, or intension of these terms and concepts doesn't involve 'activity' or causation. Keeping the history of our concepts in mind we might go so far to claim that those who deny efficacy to these properties should be considered closet eliminativists, for they deny something that is steadfastly maintained to be true. However, having said that, let me be more careful. Not all mental items have been thought to be essentially active: Berkeley takes ideas to be passive. With this in mind, it would not be conceptually outrageous to suggest that *qualia* are not essentially active, even if self, soul, pneuma and their 21st century counterparts are so considered.
- 17 For discussion of the exclusion argument see, among many others, Brian Jonathan Garrett, 'Pluralism, Causation and Overdetermination,' *Synthese* 116, 3 (1998) and 'Defending Non-Epiphenomenal Event Dualism,' *Southern Journal of Philosophy* 38, 3 (2000); Scott Sturgeon, 'Physicalism and Overdetermination,' *Mind* 107 (1998) 411-31.

If the full and complete explanation of heat involves even heat's causal effects upon conscious beings, and ignoring these effects means giving an incomplete account of heat, then causal essentialism offers a ready explanation for why, indeed, it is necessary to explain such effects for a *full* explanation of the phenomena. If causal essentialism is true, then the causal capacities of heat are essential to heat. If the causal capacities of heat are essential to heat, then any explanation that ignored one of these capacities would indeed be incomplete by virtue of ignoring an essential feature of the property, (even if that capacity involves consciousness!) Thus, dualists such as Chalmers and materialists alike, who take *physical* properties to be functional, and physical concepts to be causal role concepts, could well embrace causal essentialism as the metaphysical explanation for their semantic and explanatory considerations.

But *must* we believe causal essentialism, if we believe that physical property concepts are causal-functional and that a full explanation of the property requires an account of all its causal capacities? Perhaps not. The issue is difficult because it depends upon how we wish to construe explanation, and there is much room here to maneuver. But causal essentialism again offers us some help. If a theorist told me that she had *not* included in her explanation of a physical property an account of that property's disposition to instantiate another property, and this capacity was not itself logically derivative or dependent upon dispositions already accounted for, then indeed, I would think the explanation incomplete. But if the theorist remarks that she has accounted for all the capacities of the property in worlds with the same laws of nature as ours, but hasn't begun to give an account of the capacities of the property in worlds in which the laws of nature are radically different, I, for one, wouldn't feel let down. (How could she even begin this task? No experiments or measurements are available.)

Scientific explanation of a property demands no more than an account of the property's causal capacities in worlds with our laws of nature — a scientific explanation that omits an account of, for example, the fact that if the laws of nature were altered, light would have a capacity to travel faster, is not thereby incomplete. We could not credibly criticize a theory with the following: 'look, your theory is radically incomplete because although you explain how instances of P have the causal capacity to bring about instances of P* in nomologically accessible worlds, you haven't a clue about how and why P-instances *can't* bring about P*-instances and does bring about P**-instances in worlds where our laws are radically different.'¹⁸ A theory of physical properties

18 I take it that this logical possibility will not be easily denied, (unless we are causal

that omits explanations of capacities a property has in worlds which are nomologically *inaccessible*, is not a theory that offers incomplete explanations of these properties. But why would that be? That is, why would ‘complete explanation’ be bound to account for the capacities of a property only in the nomologically accessible worlds?

A plausible metaphysical answer is causal essentialism. The alleged causal capacities of physical properties in nomologically inaccessible worlds aren’t relevant to a complete explanation of a physical property because they are illusory. Causal essentialism takes the laws of nature to be metaphysically necessary — no actual physical property can be found in worlds lacking the actual-world laws. The laws are metaphysically necessary because they are descriptions of the causal capacities that constitute the property under investigation. So complete explanation rests with an account of the capacities of the property in nomologically possible worlds alone. Causal essentialism offers a way of understanding why scientific explanation of physical properties’ capacities need not attempt explanations of those properties’ capacities in worlds other than the nomologically accessible.

Might there be other explanations for why scientific explanation requires only an account of nomologically possible capacities? Perhaps. We might note, for example, that explanation in science has pragmatic concerns and goals — perhaps scientific explanation must contribute to, or reveal, the possibility of manipulating nature (knowledge is power), and the merely logically, or metaphysically possible (yet nomologically impossible) doesn’t contribute to power or prediction. Thus scientific explanation is limited to accounting for only the natural capacities of properties, *not* because this is all that is required for a *truly complete* explanation of those properties, but because this is all that is required of a complete *scientific* explanation. On this account, we may maintain that it is logically possible for the property to have capacities not included in the set of those it has in worlds nomologically accessible to the actual world, but note that these possibilities for the property aren’t interesting to science. But on this account, a complete *scientific* explanation of, for example electrons, is not a *complete full* explanation of electrons! The metaphysical possibilities regarding alternate laws of nature for electrons are ignored, therefore empirical science, under

essentialists.) Why couldn’t there be worlds in which P is nomologically incompatible with P* even though in worlds with our distinct laws P* depends on P. David Lewis holds to recombination, i.e. that any combination of a possible with another yields a further possibility, and in his Humean metaphysics this should be acceptable.

this conception, could not give truly full and complete account of the physical world.¹⁹

Scientific essentialism, by contrast, agrees that only ‘natural’ possibilities are relevant to the scientific explanation of a physical property. But scientific essentialism entails that physical properties lack these alleged contra-legal possibilities. The laws of nature, being descriptions of essential dispositional properties, logically supervene upon these properties, so no world can contain electrons without obeying the actual world’s laws. Science is free to ignore what isn’t real. This leaves us with the possibility that science does give exhaustive or complete explanations of the empirical properties when our science is complete according to its own aims and ideals.

So far then, I have noted that causal essentialism has many defenders and that as a metaphysical thesis it serves to help explain the nature of laws and properties. I argued that physical properties and physical property concepts are causal, and that this could be nicely explained by causal essentialism. Causal essentialism allows scientific explanation to be complete, for the description and explanation of a property’s actual-world capacities is all there is to that property, hence all there is to explain.²⁰

III Causal Essentialism and Zombie Worlds

It is perhaps clear by now why causal essentialism is incompatible with zombie worlds. Return to the description above:

w1	Fa and Gb For all x [F(x) -> there is a y, G(y)]
w2 (actual)	Fa, Ca and Gb For all x [F(x) -> there is a y, G(y)] For all x [F(x) -> C(x)]

19 It may be that science cannot give a full account of all the world’s properties because some of those properties are not epistemically accessible to human minds. I believe that is a plausible epistemological claim — but the claim discussed above is that science gives incomplete explanations of real properties even though we have no epistemic barriers like that to overcome in our science.

20 It should be noted that I am not committed to the view that humans can know everything about nature. The point is, rather, this: science’s goal is to describe and explain all the capacities of a property to effect change. Whether all these capacities are epistemically accessible given our mental capacities is an open question. The debate above is over a science that did succeed. Would it be complete?

The actual world contains a law linking the physical property instance F to consciousness C. If F is to be instantiated then F's causal and modal capacities must also be instantiated and these are described by the laws governing F in the actual world. But that means that in any other possibility in which we conceive of F we must conceive it having the capacity the laws describe. So in all other possible worlds, an instance of F must also have the capacity to instantiate C. There can be no zombie world after-all.

- [1] Causal essentialism: all the actual-world capacities of a physical property are essential to that property.
- [2] In the actual world an F-instance has the capacity to instantiate C²¹ and C is instantiated.

Therefore

- [3] In *all* worlds in which F is instantiated, F-instances have the capacity to instantiate C (and C is instantiated.)²²
- [4] The zombie world is a world in which F is instantiated without C, and the zombie world is a world lacking the law For all x [F(x) -> C(x)] i.e., in the zombie world F-instances lack the capacity to instantiate C.

Therefore

- [5]: [3] contradicts [4]. If causal essentialism is true, the zombie world is impossible.

Since [3] and [4] contradict one another, we must choose between CE and the logical possibility of zombie worlds. The theoretical power of

21 When a property's capacity is triggered resulting in the realization of a further property this will be limited to certain contexts or situations. Thus salt has the capacity to dissolve in water, and salt immersed in water will dissolve in certain circumstance, e.g. if the water is not already saturated with salt. The zombie world is one which is allegedly physically indiscernible from the actual world, so the contexts are intended to be identical. Thus I ignore the complications regarding contexts that one would normally mention when discussing capacities and dispositions.

22 Again, there are no further physical conditions to block the exercise of the causal capacity of F, because we are looking at worlds physically indiscernible with the actual world in which we assumed that F has been triggered.

CE is considerable, as I have already mentioned, and so our allegiance to this theory outweighs, I believe, any commitment we might have to this curious zombie world ‘possibility.’ Causal essentialism is a far more powerful metaphysical thesis than the thesis that zombie worlds are logically possible, and if we were forced to choose, other things being equal, the zombie possibility ought to be dropped. Naturally, to be convinced that the choice to drop a *prima facie* possibility due to the power of the alternate metaphysical view requires us to be confident that the metaphysical thesis is indeed attractive. This can’t be done in a short paper — but that we are forced to choose is interesting, and bad news for the supporters of the zombie worlds.

A Quinean methodological point is worth making at this stage of my discussion. Our allegiance to the possibility of zombie worlds is defeasible — the intuition cannot stand alone and gain its plausibility without seeing how the alleged possibility features into, supports or contradicts other commitments, even metaphysical commitments — especially metaphysical commitments! Those who think the zombie worlds are possible have not committed themselves to one intuition or judgment, but to many. Causal essentialism, for example, has to be rejected to maintain the possibility of zombie worlds.

IV Question-begging and Causal Essentialism

A common first response to this argument is to note that we may maintain the spirit of causal essentialism for physical properties but limit what is essential for those properties to their relations with other physical properties. What makes a physical property physical, goes the objection, is its essential capacities to bring about other physical properties, but not its capacities to bring about, say, phenomenal properties.²³ Put otherwise, we might say that physical properties are governed essentially by physical laws, but not by psycho-physical laws. In this way, we might deny that all the capacities of a property are essential to them. Physical properties, on this view, are essentially interrelated such that if one is realized, all will be, in accordance with their initial distribution and capacities, as described by the laws of nature. But making relations to phenomenal properties essential to the physical seems perverse and unnecessary to an adequate conception of the physical. Thus the impossibility of zombie worlds hasn’t been established.

23 Thanks to Torin Alter for this objection to an older draft of this paper.

Interestingly, we saw passages from Chalmers's work that conflict with this suggestion, so I believe Chalmers himself would not reply in this way. Chalmers did say that to *fully* explain heat we need to explain heat's relation to the feeling of heat. So Chalmers would have difficulty, it seems, in denying that the causal disposition of physical properties to bring about phenomenal properties, are essential to physical properties. To give a complete explanation of the physical property of heat did appear to involve its dispositions to cause even the sensation of heat. On this point, I think we should follow Chalmers.

In any case, the reply seems to beg the question. In a dispute over whether the phenomenal is or is not physical, we cannot claim that the physical is *defined* to exclude the phenomenal! That's winning by definition and in that case there should be no dispute at all. If we simply define the physical as excluding the phenomenal, then there is no reason to argue, or to pursue empirical investigation, on whether the phenomenal is physical; rather, we just need to check our definitions. (Naturally there would then be considerable disagreement over the existence of the phenomenal or physical, as so conceived.) But the debate cannot be treated as resulting from one party simply mistaking the meaning of a term. If any definition of the physical is to be pursued, it has to be defended, so the dualist must show that the most plausible conception of the *physical* is one that doesn't include essential dispositions to cause phenomenal states. But to show that without assuming the truth of dualism won't be easy. Note that my, admittedly brief defense of causal essentialism, involves showing the power of the view to explain puzzles that have little to do with dualism, e.g. its ability to explain what the truth-makers or truth-conditions of laws of nature are and why physical property-concepts are functionalizable. In that sense, my defense of CE does not beg any questions, yet CE is not consistent with zombie-world possibilities.

Perhaps this objection can be re-tooled to avoid being question-begging. We could leave it an open empirical and philosophical question whether the dispositions to bring about phenomenal properties or states are essential or not to physical properties. Let us not define the physical as excluding its relations to the phenomenal, but leave this an open question. However, if we maintain, that physical properties are essentially related to *less than all* the properties they are disposed to bring about then CE doesn't exclude zombie world possibilities. Call the thesis that *all* dispositions are essential 'Strict Causal essentialism' and the view that less than all are essential 'non-strict causal essentialism.' Zombie worlds would then be consistent with the modified non-strict CE. Furthermore, any defense of strict CE could, surely, be used to defend the modified version, and if so, then non-strict CE is as plausible as the strict version. So we have the following modified Causal Essen-

tialism: Physical properties have 100-n % of their causal dispositions essentially. Different values for n give us different metaphysical theses, all of which are consistent with the zombie world. So although strict CE may be inconsistent with the zombie world possibility, any less-than-strict version is consistent with the zombie world.

This objection is much harder to reply to, however, it has an ad hoc ring to it. Note that this objection doesn't object to the argument presented in the section above, so it leaves us with the conditional result: 'if *strict* causal essentialism is true, then zombie worlds are not possible.' That's still interesting, I think, and is admittedly the central yet modest point of the paper. So I suggest the debate has shifted somewhat, and we are asking whether strict or non-strict causal essentialism is in fact *true*, or which is the most *plausible*. Note we already know that rejecting strict causal essentialism is one way to preserve the zombie possibility — Humeanism is the natural starting point for many. The objection is now perhaps this: non-strict essentialism is compatible with the zombie worlds and non-strict essentialism is as reasonable to those who support the arguments for strict essentialism.

Now the best way to counter this objection is to run through the arguments for strict essentialism and to see whether they equally support a non-strict version. But that can't be done here. Rather, one can reply here that a non-strict version is, other things being equal, not as attractive as the strict version. It is important to note that the question is about the capacities of physical property instances to bring about instances of other properties (physical or otherwise). So why would one be motivated to think that a basic or fundamental physical property has several dispositions but not all are essential, while some are essential? Why would some dispositions be essential and others not?

Assume that an F-instance is causally disposed to bring about instances of G and also of H. It seems strange to say that the F-instance is essentially disposed to bring about H but not G²⁴. It is strange, because in both relations we are dealing with fundamental properties and their dispositions, and there is nothing further about F-instances or their contexts which could explain why one disposition is essential and the other is not. That means that any claim that some are essential and some are not, will be inexplicable by reference to F. Either all the relations are contingent or none are, but to have some essential and others not, would appear to involve us in an additional inexplicable fact about

24 Believing in essential properties, whether essential dispositional properties or not, is controversial and I offer no defense here for assuming that there are such properties. The issue here is whether there is a reason to distinguish, for the basic properties, causal capacities that are essential from those that are not.

F that strict essentialism is not committed to. The situation may be obscured if we take composite objects as our model. Composites, like trees or people, have inessential parts; therefore, they have at least one inessential property. Perhaps we can partially explain the latter with the former. But the basic physical properties and individuals have no further parts, or decomposition, nor are they reducible to anything further. That is, modified CE is not, in fact, as plausible as strict CE, since the modification would add a mysterious distinction to our metaphysics of nature. Again, we are not disputing whether strict essentialism is incompatible with the zombie worlds but whether a non-strict essentialism could do the job that strict essentialism does, and whether it would be equally plausible as a strict essentialism. But this is unlikely: for if instances of G lawfully brings about instances of H and no further underlying fact explains that relation, then it is unclear why another basic lawful connection does not also have the same modal properties as the first. We need a novel reason to show that the additional distinction is doing essential, and not ad hoc, work.

V Causal Essentialism, Zombie-worlds and Property Dualism

The last objection I'll consider raises an important point about causal essentialism. Surely any metaphysical thesis that itself rules out the possibility of dualism is already implausible. That is, a good argument regarding a metaphysical thesis involves making the thesis acceptable to as many people with differing assumptions, as one can. But strict CE is incompatible with dualism, which is a live metaphysical option, so the plausibility of strict CE is seriously threatened by its controversial consequences.

Modal or causal essentialism, although incompatible with the existence of zombie worlds, is not itself incompatible with a non-reductivist 'dualist' ontology regarding consciousness, or any other property, for that matter. We should not confuse the debate over whether causal essentialism regarding the physical allows for zombie worlds, with the debate over whether causal essentialism is committed to the identity of consciousness with some physical property. Although the existence of zombie worlds would be incompatible with an identity of consciousness with some physical property, and would be incompatible with modal essentialism regarding the physical, modal essentialism does not itself imply identity. But modal essentialism does rule out the possibility of a zombie world, meanwhile leaving it open whether consciousness is physical or one of many basic properties that are essentially interrelated. In a way, the concern is over a form of holism and nonidentity; the causal essentialist can appeal to a metaphysical view that commits

one to believing that there are essential dependencies between instances of dispositional properties, their triggering circumstances and their effects, that are, nevertheless, essential relations between ontologically distinct properties.

So this objection confuses the concern for dualism with the concern for the zombie-world possibilities. Zombie-world possibilities are being used as a premise in the argument for dualism mentioned above. Causal essentialism does not by itself make dualism — the view that mental properties are irreducible or non-identical with a number of other properties dubbed ‘physical’ — impossible or false. It doesn’t beg that question at all. All that causal essentialism rules out is the existence of a zombie-world possibility. That means that we lose a premise in one argument for dualism, but there are other arguments for property dualism that may be untouched by strict causal or modal essentialism.

If causal essentialism is true we may still think of mental properties as irreducible. What causal essentialism requires is that the basic fundamental properties of the world have their dispositions to produce or realize other fundamental properties essentially, such that there can be no instantiation of these properties without their dispositions to produce other properties, and when the conditions under which that disposition would be triggered are met, then a further fundamental property will, of necessity, be instantiated. Nothing in this doctrine declares that the fundamental properties (those that cannot be further reduced or instances reductively explained) are not phenomenal properties. If we accept this kind of holism, implying that the instantiation of one fundamental property requires the instantiation of others, (given their triggering circumstances), we could still agree that one of these fundamentals is phenomenal. We would need a good reason to think of consciousness as irreducible and essentially interdependent, just as the fundamental physical properties are irreducible to one another yet essentially interdependent. That is, we would need to show that there are no further mechanisms underlying the phenomenal properties that capture all the features we are certain phenomenal properties possess. But zombie possibilities cannot be appealed to here to show that the reductive explanation of phenomenal instances is incomplete.

So how likely is a non-reductive causal essentialist view of phenomenal properties? Given that, as a matter of empirical observation, phenomenal properties are only instantiated in macro-physical systems, which do have underlying mechanisms, it is unlikely that consciousness will be one of those fundamental properties. Unlikely, but not impossible. No argument can appeal to zombie-world possibilities to make the case that phenomenal properties ought to be included among these fundamentally irreducible properties, if we follow the metaphysics of causal essentialism. We would need to show, with a distinct argument,

that the mechanisms that appear to underlie phenomenal property instances are insufficient to capture all that is true of the phenomenal properties, if we have a strong and less partial case for property dualism.

In summary, if causal essentialism is true then zombie worlds are not logically possible. But causal essentialism alone does not show that phenomenal properties are physical. We can reject the zombie argument against physicalism with the resources of causal essentialism, but a causal essentialist might claim that among the fundamental properties so-related is one that ought to be dubbed non-physical. It is logically possible that the phenomenal is not physical despite the logical impossibility of zombies, given causal essentialism.

Received: August 2006