Philosophers' Imprint

VOLUME 12, NO. 13
JULY 2012

Functionalism and the Metaphysics of Causal Exclusion

David Yates

Kings College, London

1. Introduction

Take functionalism to be the thesis that mental property M is the property of having some other property that plays a certain characteristic causal role R.[1] Functionalists are usually physicalists, and so take mental properties to be physically realized, such that for any mental property M, there's a physical property P that fills R. Causal exclusion looms. Functionalism takes mental properties to be characterised by causal roles that are filled by something else. As M's realizer, P must do all the causal work (whatever that is) that R involves, apparently leaving nothing for M to do. If mental properties are causally redundant, and if causal novelty is necessary for robust ontological commitment, then mental properties aren't really real.[2] This would be no small irony if true, because historically, functionalism was motivated by the need to explain how physically different creatures could be in the same mental state; small comfort to be told that it's by dint of falling under predicates that don't pick out genuine properties. Label properties that do causal work "causally efficacious", setting aside for now the question of what causal work is. Those who argue that functionalism is consistent with the causal efficacy of mental properties typically adopt one of the following strategies: (i) deny that causal novelty is necessary for ontological commitment and argue that functional properties can inherit the efficacy of their realizers, resulting in a kind of causal over-determination;[3] (ii) defend a difference-making theory of causation that entails that functional properties are causally novel after all.[4] Proponents of (i) defend, or at least recognise the need to defend, theories of causation that vindicate their claim that functional properties inherit efficacy from their realizers. The task then is to argue that the kind of over-determination involved isn't problematic.

1. Functionalism so construed isn't limited to mental properties. My arguments in this paper depend only on the general metaphysic outlined above.
2. Kim (1992a,b, 1998). I will fill in the preceding sketch in due course.
3. Segal and Sober (1992); Bennett (2003); Witmer (2003); Kallestrup (2006).
4. Yablo (1992); List and Menzies (2009).

Proponents of (ii) needn't worry about over-determination, because according to difference-making theories of causation, functional properties make a difference their realizers don't.

Both strategies suffer, I argue, from a failure to properly analyse the notion of causal work as applied to properties. If, as is commonly supposed, the causal work of a property consists in grounding the dispositions of its bearers, then the nature of functional properties renders it impossible for them to do the same causal work as their realizers. For related reasons, I argue, whatever the merits of the difference-making account of causation, it can't be an account of the causal work that properties do. Clarification of the notion of causal work reveals a novel solution to the exclusion problem based on the relations between dispositional properties at different levels of mechanism, which involves three central claims: (i) the causal work of properties consists in grounding dispositions, (ii) functional properties are dispositions, and (iii) the dispositions of mechanisms are grounded in the dispositions of their components. Treating functional mental properties as dispositions of components in psychological mechanisms, I argue that such properties do the causal work of grounding agent-level dispositions. These dispositions, while ultimately grounded in the physical realizers of mental properties, are indirectly so grounded, through a hierarchy of grounding relations that extends upwards, of necessity, through the mental domain.

2. The Causal Exclusion Problem and the Humean Backlash

Proponents of Humean approaches to causal exclusion suppose that the causal efficacy of properties can be captured by counterfactual or nomic relations between events, typically — but not necessarily — construed as instances of the target properties. In Section 2.1, I discuss Kim's "supervenience argument" against mental causation,[5] and in Section 2.2, I show how the Humean strategies outlined above are supposed to block that argument. I postpone discussion of the causal work of properties to Section 3, where I argue that the mere fact that, by Humean lights, a property-instance *causes* some effect, doesn't entail that it's an instance of a property that does causal *work* in relation to that effect. Kim treats causation as a relation between fine-grained events, construed as property-instances. On this view mental events are numerically distinct from the physical events that realize them. Those who prefer coarse-grained events may recast the arguments that follow in terms of single events having both mental and physical properties, with the former supervening on the latter.

2.1 Kim's supervenience argument.

The most complete presentation of Kim's supervenience argument occurs in his (2005).[6] Kim offers two versions of the argument; for brevity I focus on the simplest, which concludes that no property that's not identical to some physical property could cause an instance of a supervenient property. It's the supervenience of the effect property-instance on a physical base that drives the simple version of Kim's argument.[7] Kim begins by focusing on "mental-to-mental" causation, which I take to involve instances of intentional mental properties such as thirst causing instances of behavioural properties such as drinking. The argument also applies to causal chains of mental property-instances of the first kind; I focus on causation of behaviour for reasons of exposition. Following Kim, let the mental property-instances be M and M*, and let M*'s physical base property-instance be P*. For *reductio*, suppose that M causes M*. Kim defends a principle of downwards causation, according to which the only way to cause M* is to cause P*. The idea is that since P* realizes M*, to suppose an event could cause M* without causing P* is like supposing a pill could alleviate a headache without causing any brain events.

5. Kim (1998), ch. 2; (2005) ch. 2.

6. ch. 2, pp. 39–52.

7. The more complex version depends on the supervenience of both cause and effect properties.

(DC) M causes M* by causing its physical supervenience base P*.[8]

Kim needs two further principles to show that M doesn't cause M*. First, the causal closure of the physical:

(CC) If a physical event has a [complete, sufficient] cause that occurs at t, it has a [complete, sufficient] physical cause that occurs at t.[9]

On the assumption that M causes M*, it follows from (DC) that M causes P*. But from (CC), P* must have a physical cause at the time t when M occurs. This event, P, is most naturally regarded as M's supervenience base, but the simple version doesn't depend on this. One further principle is required, the oft-cited "causal exclusion principle":

(CX) No single event can have more than one sufficient cause occurring at any given time — unless this is a genuine case of causal over-determination.[10]

Assuming that M and P both cause P*, it follows from (CX) that either M=P, or this is a genuine case of causal over-determination. Since functionalism is non-reductive, identifying M and P isn't an option. Identifying property instances entails identifying the properties they are instances of, since a property instance $(x,\mathrm{P},\Delta t)$ is identical to a property instance $(y,\mathrm{Q},\Delta t')$ if and only if $x=y$, P=Q and $\Delta t=\Delta t'$.[11] What about over-determination? Kim takes "genuine causal over-determination" to involve two independent causal chains leading to the same effect, where each would have been sufficient in the absence of the other. As he notes, that doesn't seem to be the right way to describe the way in which — putatively — M and P both cause P*.[12] In cases such as execution by firing squad, the bullets of the different marksmen over-determine the victim's death by causing instances of distinct properties in the victim. The *manner* of the victim's death is *altered* by its being over-determined in this way. This isn't to say that the victim dies a different death by virtue of this kind of over-determination — as Lewis argues, in ordinary discourse we speak of events as if robust with respect to small enough changes in the manner of their occurrence[13] — but we can at least find for each cause a different effect. A death is an instance of a highly determinable property, and one might suggest that causing a death is a matter of causing one of its determinates. If we think of determinable properties as supervenient properties, then this much follows from (DC). It's possible, then, for two marksmen to over-determine a death by each causing one of its determinates at the same time, and this is consistent with holding that the victim would have died the same death had one of the marksmen missed. Over-determining causes of this kind have different causal roles. Not so in the case of mental causation, it seems: M and P putatively over-determine P* by doing *exactly the same thing*, viz. causing P*. Call this "redundant over-determination". The only remaining option is to deny the assumption we started with, viz. that M causes P*. But by (DC), if M doesn't cause P*, then it doesn't cause M* either, and there's no mental causation.

Kim doesn't have a knock-down argument against redundant over-determination, but in earlier work, he argues, based on a principle he calls "Alexander's dictum", that if mental properties have no novel causal work to do, we should eliminate them:[14]

(AD) To be real is to have causal powers.

8. Kim (2005) p. 44.

9. *Op. cit.* p. 43. I have added the parenthetical 'complete, sufficient' to Kim's formulation because (a) 'sufficient' is clearly implicit in that formulation, and (b) without 'complete' (CC) would be consistent with certain forms of emergentism, according to which mental and physical causes combine to cause P*. I return to the completeness of physical causes in (4).

10. *Op. cit.* p. 41.

11. $(x,\mathrm{P},\Delta t)$ should be interpreted as: x's having P during interval Δt.

12. *Op. cit.* pp. 46–52.

13. Lewis (1986).

14. Kim (1992a,b).

According to (AD), there is no such thing as an entity that doesn't cause (or contribute to causing) anything. The principle may seem too strong, since it rules out the existence of abstract entities such as numbers, sets, and so forth. I am inclined to think a reasonably circumscribed (AD) can avoid such problems: we might, for instance, limit its scope to natural entities. It's one thing to claim that some entities have no causal powers, quite another to claim that mental properties have none. I needn't worry about precisely how to formulate (AD), however, since my aim in this paper is to show that mental properties have novel causal powers. My strategy will be to grant the (perhaps implausibly) strong version of (AD) above and show that even this principle does not threaten the existence of mental properties. Kim thinks that given (AD), their irreducibility implies that "mental properties bring with them … causal powers … that no underlying physical-biological processes can deliver … . [T]o be real, new and irreducible … must be to have new, irreducible causal powers."[15]

If cogent, this line of reasoning rules out ontological commitment to properties that are never more than over-determining causes. If mental properties are irreducible and real, then their causal powers must be irreducible as well, which is to say they must do something their realizers don't. According to functionalism, mental properties are defined by roles that are filled by physical properties, so a novel causal role seems out of the question.

The upshot of the argument is that there are no functional properties. The only properties that survive the cull are the properties of fundamental physics, and any properties that are identical to particular configurations of those properties.[16] The property of being a hydrogen atom survives, as does the property of being H_2O; but the property of being a neuron doesn't, unless there's some incredibly complex structural property definable in terms of properties of fundamental physics that all neurons have in common. More obviously, mental properties are eliminated, because it's hugely implausible that all those who like roast chestnuts share a physical property. Kim's causal exclusion argument, then, is a combination of two separate arguments: (i) an argument that mental property-instances are at best redundant over-determining causes; (ii) an argument that real, irreducible properties aren't redundant, so that mental properties are either unreal or reducible. Those who endorse over-determination as a response to the exclusion argument accept (i) but take issue with (ii); those who endorse difference-making causation can ignore (ii), because they reject (i).

15. Kim (1992a).

16. See Kim (2003), where he argues against Block (2003) that causal powers don't "drain away" if there's no fundamental level. This is because Kim's target is multiply realizable properties, and Kim thinks that when we get down sufficiently deep, there won't be multiple realization any more, but a potentially infinite descent through mereological levels with the properties of each L identical to structural properties of L-1.

2.2. Humean solutions to the exclusion problem

Humean solutions are Humean because they appeal to nomic or counterfactual relations between mental property-instances and behavioural property-instances to show that mental properties do causal work. Such strategies typically involve (a) defining causation in terms of laws or counterfactuals, and showing that according to the definition, mental property-instances are causes of certain behavioural effects; and (b) a tacit assumption that if mental property-instances are *causes* of behavioural effects, then mental properties do *causal work* in bringing those effects about. I grant both (a) and (b) for now, but in Section 3, I argue that (b) is demonstrably false on an independently motivated account of causal work. First, I'll outline the over-determination and difference-making solutions, beginning with the former.

Those who think that M and P over-determine P* take this to mean that M does causal work, inheriting some or all of the causal power that P has to cause P*. Some offer criteria of causal efficacy according to which we can show that if P causes P*, then — given the nature of the supervenience relation between M and P — M causes P*. Segal and Sober, for instance, think that the (possibly non-strict) law of nature that M-instances are followed by M*-instances, together

with sufficiently tight supervenience relations between M and P, and between M* and P*, is sufficient for M to cause M*.[17] This theory is supposed to account for the causal efficacy of mental properties, and so is intended to be a sufficient condition for a property to do causal work. M inherits causal efficacy with respect to M* from P's efficacy with respect to P*, via the supervenience of M and M* on P and P*, respectively, together with a law relating M to M*. While it is desirable to defend a theory of causation according to which M causes M*, most over-determinationists focus on showing that the kind of over-determination involved in M causing M* wouldn't be problematic. Kallestrup thinks of over-determination as follows:

> (OD) E is over-determined by C1 and C2 iff (i) C1 is sufficient for E, (ii) C2 is sufficient for E, (iii) if C1 had occurred without C2, E would have occurred, and (iv) if C2 had occurred without C1, E would have occurred.[18]

In firing-squad cases, all four conjuncts are non-vacuously true. There's nothing problematic about this, because we know why these conjuncts are true in such cases: those responsible for firing squads act precisely so as to make them true. But the right-hand side of (OD) is also non-vacuously satisfied in some cases without there being any explanation of why this is so. It's possible that two assassins independently decide to assassinate the same person at the same time without a common cause that might explain this. This too is unproblematic, provided it doesn't happen very often. If it did, there would be widespread inexplicable coincidence of forward-looking causal powers, which — grant for the sake of argument — would be a very bad thing. Now M and P satisfy (OD) with respect to P*, but in a different way. Let P=C1, M=C2, P*=E. Since P is M's supervenience base, it is plausibly metaphysically necessary that if anything has P, it has M.[19] If there are no P-occurs-without-M worlds, conjunct (iii) is vacuously true.

The fact that M supervenes with metaphysical necessity on P also — on the assumption that M inherits P's causal efficacy — renders it unmysterious that there are two causes of P*. If widespread coincidence is the reason over-determination is problematic, not only is the kind of over-determination we have here unproblematic; we also have a test — non-vacuous truth of all of (i)–(iv) — for the potentially problematic kind. Bennett takes a slightly different view, arguing that the non-vacuous truth of (i)–(iv) is necessary for over-determination *simpliciter*, meaning that M and P don't over-determine P* at all.[20] The main import of this difference relates to (CX). Bennett must reject (CX) — M and P are distinct causes of, but don't over-determine, P*. Kallestrup, on the other hand, can hold that M and P do over-determine P*, but in a way we shouldn't worry about. It matters little for my purposes precisely what 'over-determination' means.

Assuming M inherits its causal powers from P, it remains only for over-determinationists to rebut Kim's redundancy argument. According to (AD), nothing lacks causal powers. Functional properties are irreducible to their bases. Crucially, Kim takes (AD) to imply:

> AD*: To be real and irreducible is to have irreducible causal powers.

But as others have pointed out, (AD) does not imply (AD*).[21] From (AD), it follows only that if mental properties are real and irreducible, then they have causal powers and are irreducible. As Kallestrup argues, there's no reason to take 'irreducible' to qualify the causal powers of

17. Segal and Sober (1992). Segal and Sober think in terms of Davidson-events having both mental and physical properties, and supply a sufficient condition for it to be in virtue of M that the P∧M event causes the P*∧M* event. I have recast their theory in terms of Kim-events for consistency. See also Witmer (2003), esp. pp. 205 ff.

18. Kallestrup (2006), p. 471. Adjusted for typographic consistency.

19. Kim (2005) attempts to block the argument that follows at this stage, arguing that M's supervenience on P is nomologically necessary and backed by bridge-laws. I don't find this persuasive, but haven't the space to take up the issue here.

20. Bennett (2003).

21. Stephan (1997); Kallestrup (2006).

properties rather than the properties themselves, unless of course we take the view that properties are exhausted by their causal powers. In that case, the only thing that *could* make a supervenient property irreducible is bestowing causal powers its base property doesn't.[22] Over-determinationists can reply that there are other ways for a supervenient property to secure irreducibility to its base. Provided M can inherit the causal powers of P, and M has a kind of novelty that doesn't require it to bestow novel causal powers, then there's no pressure to eliminate. We might, for instance, endorse Shoemaker's subset theory of realization, according to which M supervenes on P because its causal powers are a proper subset of the powers of P. This would seem to preclude our identifying M and P, but whether it's enough to secure the kind of novelty over-determinationists need is a moot point.[23]

Those who endorse difference-making causation as a response to the exclusion problem follow Yablo in thinking that causes must be *proportional* to their effects.[24] M* has many distinct possible supervenience bases, of which P* is one; similarly, *mutatis mutandis*, for M and P. The idea, informally, is that P isn't proportional to M*, because P is causally sufficient for a specific realization of M*, viz. P*. Since M* might have occurred differently, P causally explains why M* happened in a particular way, but not why it happened *simpliciter*. Similarly, there will be property-instances that fail to explain M* due to not being specific enough: the property of having some mental property, for instance. The cause of M*, by contrast, ought to be a property-instance that's *just right*, in the sense that it causes M* however it or M* are realized, and M fits the bill. Yablo's notion of proportionality is typically now described in terms of difference-making. Following List and Menzies, define difference-making as follows, where F and G are property-instances:[25]

DM: F makes a difference to G iff: (i) F occurs $\Box\!\!\rightarrow$ G occurs, and (ii) $\neg$(F occurs) $\Box\!\!\rightarrow$ $\neg$(G occurs)

According to (DM), F makes a difference to G if and only if the set of closest F worlds to actuality are G worlds, and the set of closest non-F worlds are non-G worlds. Now let's evaluate the following counterfactuals:

(i) M occurs $\Box\!\!\rightarrow$ M* occurs

(ii) $\neg$ (M occurs) $\Box\!\!\rightarrow$ $\neg$(M* occurs)

(iii) P occurs $\Box\!\!\rightarrow$ M* occurs

(iv) $\neg$(P occurs) $\Box\!\!\rightarrow$ $\neg$(M* occurs)

It's easy to see that (i), (ii), and (iii) are true. The closest M worlds to actuality will be worlds in which M is realized by some physical property that causes a realizer of M*; the closest non-M worlds will be worlds where it has no physical realizer property, and at which no realizer of M* occurs; and the closest P worlds will be worlds at which P* occurs. However, the closest non-P worlds to actuality will — so the argument goes — be worlds at which some other realizer of M occurs. Intuitively, this seems correct: a world at which an alternative realizer P′ of M occurs, which is similar but not identical to P, and at which an alternative realizer P*′ of M* occurs, is closer than worlds at which neither M nor M* occurs. But if this is so, then (iv) is false, and P doesn't make a difference to M*.

This kind of argument holds wherever causation of a supervenient property-instance is involved, provided the occurrence of that property-instance is insensitive to the precise manner of its realization. Interestingly, such cases don't falsify (CX), but that's because P fails to be a cause of M*. This is a case of M's difference-making causal

22. Kallestrup (2006) pp. 468–470

23. Shoemaker (2001). Kim (2010) suggests that Shoemaker's theory is in fact a form of type-identity theory, but these matters are beyond the scope of this paper. I will note, however, that if P has as a constituent a physical property P′ whose powers are the same proper subset of P's powers that M inherits, then there's nothing to prevent identification of M with P′.

24. Yablo (1992).

25. List and Menzies (2009); Menzies (2008).

role with respect to M* excluding a similar difference-making role for P. What's gone wrong with the exclusion argument, then? It's not obvious what we should say. If we substitute 'difference-making cause' for 'cause' in (CC), then if we treat M* as a physical event (*qua* behavioural) then (CC) comes out false, because M* doesn't have a physical difference-maker. If we substitute difference-making causation in (DC), it too comes out false. M isn't a difference making cause of P*, since ¬(M occurs □→ P* occurs). Perhaps we could then keep hold of (CC) by employing a more flexible notion of causation. There is, after all, a sense in which M* does have a sufficient physical cause. P is sufficient for P*, and the supervenience relation between P* and M* is synchronic and non-causal, so it would seem foolish to deny that *some* kind of causal relation holds between P and M*, and I see no reason why this causal relation shouldn't satisfy proponents of (CC).[26] Let's say that difference-making causation involves a rejection of (DC): M causes M*, but not by causing P*, so there isn't any over-determination, and M's causal role is secure.

But doesn't P still do all the causal work involved in M*'s occurrence? Well, not if causal work is understood as making a difference! It's a fact about the world we live in that even though P* has a sufficient physical cause P, it doesn't follow that P does all the causal work involved in M*'s occurrence, even though the relation between P* and M* is non-causal. Since P isn't a difference-making cause of M*, M has novel causal power: it makes a difference to M* that nothing else does. At this point one might suspect, with Kim, that difference-making causation is a cheat:[27] the theory proposes two independent levels of causal work, M making a difference to M* and P making a difference to P*. What's odd about this is that P causally necessitates M*, in the sense that it causes something, P*, that non-causally necessitates M*. If we think of the causal work required for M* to happen in terms of the kind of work a builder has to do to build a house — pushing, lugging, chopping — then how can there be any left for M to do, given P's sufficiency for P*? But that, say Humeans, is the wrong way to think about causal work. Builders certainly do this kind of work when they build houses, but that doesn't mean that events do it in causing other events. As Sider points out, it's surely wrong to think of causal work as "a kind of fluid divided among the potential causes of an effect", such that "[i]f one potential cause acts to produce an effect, that fluid is used up, and no other potential cause can act".[28] If that's how Kim thinks of causal work, then the burden of proof is on him to show that this way of thinking is more plausible than the difference-making theory — no easy task.

Let's take stock. There are two solutions under consideration. The first says that M and P both cause M*, and goes on to say either that this isn't a genuine form of over-determination, so that (CX) is false, or that it is, but of a non-problematic kind. Still, M shouldn't be eliminated, because it can earn its ontological keep without being causally novel. The second says that M is causally novel, because on an independently motivated theory of causation, M and not P causes M*, and it does so without causing P*, so (DC) is false. Both strategies assume that if we can show that an appropriate Humean relation holds between M and M*, then we will have shown that M does causal work. And that's where they go wrong.

3. Causal work and the inadequacy of the Humean response

Humeans assume that if an instance of mental property M causes an instance of behavioural property B, then M is causally *efficacious* with respect to B, in the sense that it does the kind of causal work relating to B's occurrence that exclusionists say it doesn't do. I'll now argue that this seemingly innocuous assumption is inconsistent with an independently plausible and widely held view concerning the kind of causal work that properties do. Most philosophers who think properties do causal work think that this work consists in

26. Yablo calls this relation "causal sufficiency", and distinguishes this from causation, with the latter requiring proportionality of cause and effect.

27. Kim (1998), ch. 3.

28. Sider (2003), p. 721.

grounding the dispositions of their bearers.[29] This is a commonplace, although sometimes stated as the claim that properties bestow causal powers.[30] The central idea is that things have dispositions to issue in certain types of effects when appropriately stimulated, and that these dispositions are grounded in their intrinsic properties. By grounding the dispositions of particulars, efficacious properties thereby ground causal relations involving those particulars.

The reader may worry that this understanding of causal work isn't one that Humeans would accept, so a little clarification of my aims is in order before proceeding. I think there are good independent grounds for understanding causal work in terms of grounding dispositions, and that this makes good sense of what exclusionists have in mind when they say there isn't any causal work left for physically realized functional properties to do. Thinking of causal work this way makes for persuasive arguments that neither the over-determination nor difference-making strategies work as responses to the exclusion problem. However, I don't appeal to causal work as grounding to argue against broadly Humean approaches to mental *causation*. On the contrary, what I propose to do is: (i) grant the exclusionists an independently plausible notion of causal work that shows why neither the over-determination nor difference-making strategy solves the exclusion problem; (ii) grant Humeans the difference-making theory of event causation; and (iii) show that given difference-making causation, there's plenty of novel causal work for functional properties to do, so that Humeans needn't reject the view that properties do causal work by grounding causal powers.

29. The claim that dispositions are grounded doesn't entail that they have *categorical* grounds. My position is consistent with dispositional essentialism about fundamental properties, which I can understand as the claim that such properties essentially ground certain dispositions. See Shoemaker (1980); Bird (2007). I can also treat fundamental properties as pure, *un*grounded powers, as in Molnar (2003), provided such powers aren't identified with dispositions in my sense, and so can be thought of as grounding them.

30. Shoemaker (1980); Wilson (2002); McLaughlin (2006).

3.1. Dispositions

Famously, sentences of the form '*x* is disposed to M when C', can't be analysed in terms of subjunctives of the form 'were C to occur, *x* would M'.[31] Suppose a vase is disposed to shatter when struck by an object with momentum ≥ *m*, and label this disposition D_S. (Hereafter I sometimes omit 'by an object with momentum ≥ *m*' for brevity.) Suppose the vase has D_S iff, were the vase to be struck, it would shatter. Finks falsify the putative analysis in both directions. *This* vase is Gandalf's favourite, and he'll intervene, whenever it's struck by an object whose impact would otherwise have shattered it, casting a spell to alter its atomic structure so that it doesn't shatter. Assuming dispositions are intrinsic properties, the vase has D_S prior to being struck, but the analysans isn't satisfied. Lewis proposes that the simple conditional analysis can be fixed by modifying the analysans to 'the vase has an intrinsic property P such that were it to be struck and retain P, it would shatter because of its having P and being struck'.[32] Gandalf's favourite vase satisfies this analysans, so is *disposed* to shatter when struck, even though it doesn't. However, masks and antidotes falsify the revised analysis in the left-to-right direction.[33] Fragile vases can be wrapped in protective packaging so that they don't shatter when struck. Such vases retain all their intrinsic properties, and ought therefore to retain their dispositions, but Lewis' analysis entails that such vases don't have D_S. Poisonous substances are disposed to kill when ingested, but don't cause death when taken with their antidotes. But antidotes don't alter the intrinsic properties of poisons, which ought therefore to retain their disposition to kill when ingested.

It's difficult to see how to rule out finks and antidotes in a principled way. A mask or antidote for one disposition needn't work on another. For this reason, modifications that appeal to normal or ideal conditions, or (if different) *ceteris paribus* clauses, tend towards vacuity: 'the vase has D_S iff were it struck in ideal conditions, then it

31. Martin (1994).

32. Lewis (1997); simplified for exposition.

33. Johnston (1992); Bird (1998).

would break' isn't terribly informative if ideal conditions can only be specified in terms of their enabling the manifestation of D_S.[34] A planet would shatter when struck in those conditions in which it would shatter when struck — even if no such conditions are possible. All of which is bothersome for me, since the central argument of this paper depends on construing causal work as the grounding of dispositions. Without saying something about dispositions, it won't be clear how properties ground them, or why we should call it causal work when they do. So here goes:[35]

> (DISP) *x* is disposed to M when C iff *x* has an intrinsic property P in virtue of which [there is a set of nomically possible background conditions $\{B_1, \ldots, B_k\}$ such that if *x* were in any of the B_i and C occurred, *x* would M.]

(DISP) says nothing about the nature of the conditions in which *x* would M if C, except that they are nomically possible. Treat *x*'s being in such conditions as a relational property of *x*. I say that the intrinsic properties of things determine the range of conditions in which they would exhibit certain responses to certain stimuli. More formally, on my account, the grounding property P of the disposition to M when C is an intrinsic physical property in virtue of which: $\exists B_1, \ldots, B_k \forall i[\{B_i(x) \wedge C(x)\} \Box\!\rightarrow M(x)]$, where $1 \leq i \leq k$. (DISP) handles finks and masks. Assume for the sake of argument that wizards are nomically possible.[36] No vase is necessarily Gandalf's favourite, so even a vase protected by Gandalf has intrinsic properties in virtue of which there's a range of nomically possible conditions such that if it were struck in those conditions, it would shatter. Such a vase therefore has D_S. Similarly, a carefully packaged vase needn't be so packaged, so according to (DISP) will have D_S even though the closest worlds where it's struck are ones where it doesn't break.

A *prima facie* difficulty arises with "reverse finks". Consider a cubic block of granite, and suppose Gandalf hates the block's shape so much that whenever it's struck, he changes its atomic structure so that the impact is sufficient to break it. The block has an intrinsic property — its shape — in virtue of which there's a range of conditions in which it would shatter when struck, so the stone has D_S after all. It also, however, has intrinsic properties in virtue of which there's a range of conditions in which it *wouldn't* shatter when struck — a proper subset of the possible conditions in which no wizard hates cubes. What this means is that the block is both disposed to break when struck by an object of momentum $\geq m$, and disposed to remain intact when so struck. The disposition to M when C is not, according to (DISP), a contradictory of the disposition to not-M when C — provided there's no *overlap* in the possible background conditions in which it would M when C and those in which it wouldn't. Note that this isn't the context-sensitivity others have pointed out in the satisfaction of predicates like 'fragile'.[37] A chair regarded by its Lilliputian designers as robust will be seen as fragile by Gulliver, but both can agree that it's disposed to break when Gulliver sits on it. They will simply disagree about whether this disposition ought to be counted as a case of *fragility*. That things can be disposed to shatter when struck and disposed to not-shatter when struck has nothing to do with this kind of context-sensitivity, but does point to another kind. If the background conditions in which granite blocks would shatter when struck were likely to obtain, rather than distant nomic possibilities, we might well regard such things as fragile.

3.2. Causal work as grounding dispositions

I take grounding to be a transitive, irreflexive, and asymmetric relation which holds between facts — construed as things having

34. Martin (1994); Fara (2005).

35. My central arguments will go through on other accounts of dispositions, for instance Lewis (1997); Fara (2005). In general, my account will work for any account of dispositions that explains (i) their grounding, (ii) the relationship between grounding dispositions and causality.

36. It doesn't matter if the actual laws of nature rule out wizards, because I want (DISP) to be true at worlds where the laws of nature that hold there don't.

37. Mumford (1998); Hawthorne and Manley (2005).

properties — and reflects ontological priority.[38] If an entity x having a disposition D is grounded in its having some property P, then x has D in virtue of having P, and x's having P is more fundamental than its having D. Since grounding entities are more fundamental than those they ground, nothing can be its own ground. (Fundamental entities are *un*grounded, not *self*-grounding.) Further, if P grounds D, it can't be the case that D grounds P; for otherwise each would be more fundamental than the other. So understood, grounding must be both irreflexive and asymmetric. I hold that grounding is transitive because I take *X*'s ground to be a metaphysically explanatory reason for *X*. If psychology is grounded in neuroscience, for instance, then neuroscience explains why we have psychological properties. But if psychology, in this sense, grounds economics, then I take that to imply — at least in *some* sense — that neuroscience explains economics.[39]

Intrinsic properties determine how things would respond to various stimuli in a range of nomically possible circumstances. Loosely speaking, the vase's atomic structure grounds D_S by being the kind of structure that can't absorb more than a certain amount of energy without undergoing the kind of rearrangement that counts as a shattering. Having this structure determines that there are conditions in which the vase would break when struck by objects that have at least that much energy to give. These conditions will include that nothing (such as bubble wrap) prevents the vase from absorbing the energy of the striking object; that no wizard loves the vase so much he'll change its structure so that it can absorb the energy without shattering; and so on. Since dispositions manifest reciprocally,[40] the same goes *mutatis mutandis* for the striking object — a hammer, say — whose properties will determine how much energy it can transfer, in a range of conditions, to objects it strikes. Its properties thereby determine that there are conditions in which it can shatter objects incapable of absorbing more than a certain amount of energy without shattering. It's because these two sets of conditions overlap that hammers are best kept in the shed, where there are no vases. Let's turn now to the question of what's causal about causal work as I understand it.

Ordinary causal talk enables us to identify properties that ground dispositions, and the stimulus conditions of those dispositions, as causes.[41] In 'the vase broke because of its atomic structure' we do the former; in 'the hammer's impact caused the vase to shatter', the latter. Thinking of the relata of causation as property-instances blurs this distinction: the locution 'the vase's having the property of being struck by a hammer with momentum ≥ *m*' refers not only to properties that do causal work in grounding the power of the hammer to shatter the vase, but also to the stimulus condition of that disposition — the *striking* of the vase by the hammer. Whether or not this condition is met makes a difference to whether or not the vase shatters. Since stimulus conditions don't ground the dispositions whose stimulus conditions they are, but do make a difference to their manifestations, it follows right away that there are difference-making causes that do no causal work, which should give us grounds for doubting the difference-making solution to the exclusion problem. I don't depend on this in what follows, however, and for now will focus on how grounding dispositions also grounds difference-making causal relations between stimulus conditions and manifestations.

A vase's intrinsic properties determine a range of possible conditions such that were it struck in any of those conditions, it would shatter. Suppose such a condition obtains at the actual world, and that a vase is struck and shatters. The closest possible worlds to actuality at which the vase is struck will be worlds at which the same background conditions hold: the vase isn't bubble-wrapped, and it isn't Gandalf's favourite. Clearly the vase will also have D_S at these worlds, and so shatters. Hence, at our world the vase is such that (a) had it been

38. I'll often speak of properties grounding other properties, but this is shorthand.

39. I say more about the difference in the sense of 'grounds' between (*e.g.*) 'neuroscience grounds psychology' and 'neuroscience grounds economics' in Section 4.2, where I distinguish proximal from distal grounding.

40. Heil (2005); Martin (2007).

41. We can also identify the dispositions themselves as causes. More on this in Section 3.4.

struck, it would have shattered. The vase's intrinsic properties also determine other dispositions, such as the disposition to *remain intact*, D_I, when struck by an object of momentum < *m*, or no object at all. There will be significant overlap in the background conditions for D_S and D_I; in fact, D_I was manifesting right before the vase was struck. Given that the shattered vase was in background conditions for both D_I and D_S, it follows straight away (assuming also that it's in the same conditions at the closest worlds to actuality where it *isn't* struck) that (b) had it not been struck, it wouldn't have shattered. Now given the difference-making theory of causation, it's clear how grounding an object's dispositions also grounds its causal potentialities, because (a) and (b) are jointly necessary and sufficient for the striking to be a difference-making cause of the shattering.[42]

One might object at this point that it is things, not properties, that do causal work. The hammer exerts a force on the vase, thereby transferring energy to it — an informal version of the physicist's notion of work. The exclusion problem, however, depends on there being a sense of 'causal work' that makes it true that *properties* do causal work, and transferring energy isn't it. To see this, consider that energy is itself a causally efficacious property par excellence. Gamma-rays cause skin burns and radiation sickness in virtue of their high energy, and the causal processes by which they do so involves transfer of some of that very energy to the unfortunate victim. One such process is pair production, whereby a photon of energy greater than 1.022M*e*V interacting with a heavy nucleus produces an electron-positron pair.[43] The electron, for instance, may then interact with organic tissues, ionizing them by ejecting electrons, resulting in a positive charge. My point here is that it isn't the *energy* of the photon that transfers energy to electrons in the tissues, causing their ejection. Having an energy greater than 1.022M*e*V merely *disposes* the photon to bring about pair production when appropriate conditions obtain, which in turn produces electrons that are disposed to bring about ionization of tissues when further such conditions obtain. Properties ground the dispositions of particulars, thereby determining the range of causal interactions in which they could be involved.

3.3. *Functional realization as causal work*

Functional properties are physically realized, and their realizers do all the causal work associated with their defining roles. Given that causal work is grounding dispositions, it follows that realizers ground all the dispositions particulars need in order to enter into role-defining causal relations. We might say that physical properties realize functional properties by grounding a certain characteristic set of dispositions, with realization ontologically dependent on grounding. In this section, I argue that this is double-counting, because realization and grounding are the same relation. I first argue that functional properties, as typically construed, *are* sets of dispositions, and proceed to show that given this, realizing a functional property is grounding it. Functional properties don't just *depend* upon the causal work their realizers do, *they are* (at least some of) *that work*.[44] As we'll see, this identity of realization and causal work renders the over-determination and difference-making strategies ineffectual against the exclusion argument. It also paves the way for my positive theory of mental causation, which I present in part (4). The arguments I give below depend on (DISP), although they could, I think, be adapted to other broadly Humean theories of dispositions. (DISP) suggests (although it doesn't entail) that dispositions are second-order properties; I treat them as such in what follows.

42. There will be other Humean theories of causation and dispositions that have the same explanatory virtues, so my central arguments don't depend on the present ones being correct. Perhaps the reader is prepared to accept, without a particular theory of causation or of dispositions, that what grounds *x*'s disposition to M when C thereby grounds the causal relation that obtains on some occasion when C occurs and *x* Ms.

43. Gamma photons of energy greater than $2m_e c^2$ are required (where *m* is the electronic mass, and c the speed of light), with the energy of the photon being converted, according to Einstein's energy-mass equation, into the rest masses of the electron and positron. The details needn't concern us here.

44. Many are prepared to accept without argument that functional properties are complex dispositions. See for instance McLaughlin (2006).

Using '$\lambda x.p$' to denote the property of being an x such that p, letting the existential quantifier range over intrinsic physical properties, and where $1 \leq i \leq k$, we can represent the disposition D to M when C as follows:

(1) $D = \lambda x.\exists P[P(x) \wedge \{\exists B_1, \ldots, B_k \forall i[\{B_i(x) \wedge C(x)\} \boxdot\!\!\rightarrow M(x)]$ in virtue of P$\}]$

Given (1), D is the property of having an intrinsic property P in virtue of which there exist conditions such that if the bearer were in them and subject to C, it would M. Functionalism about the mind is typically characterised by appealing to psychological theories: a given functional property is the property of having some property that occupies a causal role specified (in a way to be clarified) by that theory. In the now standard way, let's write the theory (whatever it may be) as $T(F_1, \ldots, F_n; O_1, \ldots, O_m)$, where the F-terms are the predicates that denote mental properties, and the O-terms are everything else. Replacing the F-terms with appropriately indexed variables, and prefixing the resulting formula with an existential quantifier, we get T's Ramsey sentence, R(T): $\exists P_1, \ldots, P_n[T(P_1, \ldots, P_n; O_1, \ldots, O_m)]$. Functionalists define the ith mental property F_i as follows:

(2) $F_i = \lambda x.[\exists P_1, \ldots, P_n[T(P_1, \ldots, P_n; O_1, \ldots, O_m) \wedge P_i(x)]$

Given (2), F_i is the property of having some other property that occupies a specific place in a network of states that jointly satisfy T. It isn't immediately obvious from (1) and (2) that functional properties are dispositions, but I'll now argue that on reasonable assumption, they are. I don't have anything like a proof; rather, I suggest a series of further constraints on (2) if it is to adequately define functional properties, with each one bringing it closer to (1).

In order to define a specific functional property F_i, the right-hand side of (2) must specify a causal-nomic structure and the part of it occupied by F_i's realizer P_i. For this to be the case, the O-language must contain causal-nomic predicates. This isn't a problematic requirement here, because for present purposes the definienda are mental properties, so we only need to Ramsify away the mental predicates, and the causal-nomic ones aren't mental.[45] So far, I claim only that if (2) is to define a functional property, R(T) must specify the causal structure that the properties $P_1, \ldots, P_n$ must satisfy in order to count as realizers of $F_1, \ldots, F_n$. But now if, as I am assuming, causation is difference-making, then R(T) must contain a conjunction of subjunctives specifying the differences physical properties need to make in order to count as realizers of mental properties. Suppose for the sake of argument that T is folk-psychology, and that part of what T claims about the property W of wanting some X is that given a specified psychological state F_T, wanting X causes so acting as to acquire it.[46] Given difference-making causation, this component of T can be written:

(S) $\forall x[\{[F_T(x) \wedge W(x)] \boxdot\!\!\rightarrow A(x)\} \wedge \{[F_T(x) \wedge \neg W(x)] \boxdot\!\!\rightarrow \neg A(x)\}]$

where A is the property of so acting as to acquire X. Holding F_T fixed, wanting X makes a difference to whether you try to get it. R(T) will preserve this subjunctive structure as a constraint on the P_i, so (2) defines functional properties by means of subjunctive conditionals, which is how (1) defines dispositions. My argument assumes that Ramsey sentences specify causal structure using causal *predicates*. Readers persuaded that psychological theories have an explicitly subjunctive form won't need to commit to this claim, or the difference-making theory of causation, to see this similarity between (1) and (2).

45. The reader will be reminded here of the familiar Newman objection to structural realism in the philosophy of science. Structural realists often employ Ramsey sentences to define theoretical terms, but if the non-theoretical terms left un-Ramsified consist solely of logical and observational (in the positivists' sense) vocabulary, then it's provable that the Ramsey sentence constrains only the cardinality of the theoretical domain. These matters are beyond the scope of this paper, but note that structural realists wanting to avoid the Newman objection can argue that causal vocabulary belongs in the O-language and so shouldn't be Ramsified. I needn't argue anything so strong: for my purposes it suffices that the causal vocabulary isn't in the *F*-language, *i. e.*, that it's non-mental. See Papineau (2010), pp. 381–2 for discussion.

46. F_T will be a conjunction of other mental properties such as not having stronger desires for not-*X*, believing that the actions necessary to acquire X are morally permissible, etc.

Several differences remain: (a) (2) doesn't say anything about its being *in virtue of* the P_i that the specified causal structure obtains, and makes no explicit mention of (b) stimulus conditions, or (c) background conditions. I'll now briefly address these points in turn.

First, (a). It's widely supposed in the literature that Ramsified theories can fully define physically realized functional properties, but if causation is difference-making (or more generally, Humean), then this isn't so. The reason is that implicit in the notion of realization is the claim that the physical realizers of a functional property do all the causal work associated with the role that defines it. That, as I said, is why there's a causal exclusion problem. In the next section I'll present an argument that doesn't depend on the view that functional properties are dispositions (and one that does) to the effect that making a difference (or more generally, standing in a Humean causal relation) to some effect isn't sufficient for doing *any* of the causal work involved in its occurrence. This being so, if the Ramsey sentence only places Humean causal constraints on the realizers, its truth will be consistent with those realizers not doing any of the causal work associated with the roles they realize. In addition to such causal claims, then, we need to include in (2) the further stipulation that the P_i stand in their various role-defining causal relations *in virtue of* the P_i.

Let's move on to (b). (1) characterises a disposition D in terms of a stimulus condition C and manifestation M, whereas (2) characterises a functional property F_i in terms of a network of physical states, causally related (in virtue of their physical properties) as specified in T. I claim that the F_i, so understood, are dispositions, which will sometimes manifest as behaviour, sometimes as other F_i; but where are their stimulus conditions? What makes for the *prima facie* disanalogy between (1) and (2) is that it's natural to think of hammers as having dispositions to break vases, which have reciprocal dispositions to be broken by hammers; and in addition to suppose that the stimulus conditions of these dispositions are that their bearers come into contact. Where's the psychological analogue? Suppose the stimulus condition of psychological dispositions are simply that they are properties of the same agent — then T encodes such stimulus conditions in virtue of quantifying over agents. Perhaps we can do better. Suppose I want some food, but don't yet know it — distracted by philosophical theorising, I haven't reflected on my current desires. Part of the psychological state F_T relative to which my wanting food makes a difference to my having some will be my entertaining both the desire and the rest of F_T. Perhaps jointly entertaining beliefs in a deliberative process is, in a sense, bringing them into contact with each other. If so, then T will contain terms for conditions analogous to the stimuli such as contacts, impacts, and so forth, to which we appeal when defining dispositions like fragility.

And finally, (c). I treat the realizers of the disposition to M when C as determining a range of conditions such that if their bearers were in those conditions and C happened, they would M. As we saw in Section 3.1, this enables my approach to deal with finks and masks. I argued above that if R(T) is to define a causal structure, then T must make general causal claims about mental properties, and hence, assuming difference-making causation, must imply subjunctive claims of the form given in (S). I omitted *ceteris paribus* (cp) clauses for simplicity, but it's now time to put them back in. Since the psychological domain isn't causally closed, T must contain implicit cp clauses — there *isn't* a conjunction of psychological properties such that if you have all of them, then wanting food makes the appropriate difference to eating. Suppose, for instance, that there's a drug — *backwards powder* — whose primary effect is that humans who have F_T and ingest it are such that if they wanted food, they wouldn't have any, and if they didn't, they would. We might deal with the implicit cp clauses in T by simply incorporating them into (2), qualifying the truth of the relevant subjunctives. Or, as I prefer, we could deal with them by saying that the realizer properties $P_1, \ldots, P_n$ aren't properties in virtue of which those subjunctives *hold*, but properties in virtue of which there exist a range of *conditions* in which they hold. One such condition, for us, is the absence of backwards powder, but this isn't so for all possible realizers of our mental states. Martians are a case in point: on Mars,

backwards powder is a naturally occurring mineral the ingestion or non-ingestion of which makes no difference at all to whether or not they satisfy T.

Functionalist mental properties are, I conclude, a complex kind of dispositional property. Not, it must be said, the property of having *a* property in virtue of which there are conditions in which *a certain* subjunctive is true; rather, the property of being in one of a *range* of states in virtue of which, collectively, there exist conditions in which a large *conjunction* of subjunctives is true. But they are no less dispositional for their added complexity.

3.4. Why Humean causation isn't sufficient for causal work

Although I endorse the difference-making theory of causation, finding causal work for a property to do is harder than showing that its instances are causes. I argued above that (A) the causal work of properties consists in grounding dispositions; and (B) functional properties are complex, physically realized dispositions. It follows from (A) and (B) that functional realization is grounding: functional properties are (at least some of) the causal work that their realizers do. This licenses a simple argument for the view that (C) being a Humean cause of some effect isn't sufficient for doing any of the causal work involved in its occurrence. Those who argue against the causal efficacy of dispositions, construed as second-order functional properties, simply employ the exclusion argument. Here's Jackson: [47]

> Consider … a fragile glass that shatters on being dropped because it is fragile … .There will be … a certain kind of bonding P between the glass's molecules which is responsible for the glass being such that [it breaks when dropped] … . But then it is bonding P together with the dropping that causes the breaking; there is nothing left for the dispositions itself to do. All the causal work is being done by the bonding P together with the dropping.

Why, goes the standard over-determinationist response, can't the fragility (F) do the same causal work as P? Jackson doesn't say, but given that the causal work of P consists in grounding F, the answer is clear: *because grounding is irreflexive*. It isn't just that there's no causal work *left* for F to do; there doesn't seem to be any work it could *possibly* do. We can now see why Humean theories of causation don't capture the notion of causal work at stake in the exclusion argument. Suppose, for instance, that the nomic-subsumption theory of causation is true. Dispositions supervene on their physical realizers and are plausibly lawfully related to their own manifestations, but they are patently not self-grounding. The same goes *mutatis mutandis* for any counterfactual theories, such as the difference-making theory, which entail that dispositions cause their own manifestations.[48]

Difference-making theories of causation, when taken as sufficient conditions on causal work, imply that dispositions do causal work their realizers don't. Consider the fragility of the vase F and its physical realizer P. Assuming the vase's shattering, S, is realizer-invariant, F makes a difference to S that P doesn't. The closest non-F worlds are non-S worlds, and the closest F-worlds are S-worlds, which is all that's required for F to make a difference to S. By contrast, the closest non-P worlds are worlds where some other realizer P′ of F occurs, together with a different realizer of S. Since the closest non-P worlds are S-worlds, P doesn't make a difference to S. F indexes the vase to a different portion of modal reality than P, and so makes a difference to S that P doesn't.[49] That dispositions are difference-making causes of their manifestations shows that making a difference isn't sufficient

47. Jackson (1996), p. 393. Adjusted for typographic consistency. See also Prior, Pargetter and Jackson (1982); McLaughlin (2006).

48. The reader may worry that dispositions aren't causally relevant to their manifestations, but as McKitrick (2005) argues, independently plausible theories of causal relevance imply that they are. If we treat difference-making as an account of causal relevance, it's a case in point.

49. This will prove important in Section 4.4.

for doing causal work. The property F *is* the glass's disposition to shatter when dropped, and so, since grounding is irreflexive, can't ground that disposition. Difference-making causation identifies the causal work done as a cause, which seems right. Causal work makes a difference to what happens; if it didn't, what would be the point of doing it? But this means we shouldn't expect all difference-makers to do the kind of causal work the exclusion argument denies to mental properties.[50] Conversely, not all properties that ground a disposition make a difference to its manifestations.

It will be noted that in Section 3.3, I presupposed (C) above as part of my argument for (B). But the argument for (C) just given presupposes (B), so I still owe the reader a further argument for (C) that doesn't depend on (B). I'll now argue that (C) follows from (A) alone, together with the reasonable claim that realization is a form of grounding, even if functional properties aren't dispositions. Functional property F is the property of having some property P with causal role R. For P to have causal role R, given (A), is for it to ground the dispositions that characterise R: label this set of dispositions D_i. Given that P realizes F by doing the causal work associated with role R, it follows that P realizes F by grounding D_i. It doesn't matter which P grounds D_i—it's D_i that really matters to F's realization. Since F ontologically depends on D_i, and having D_i (grounded by some P) is sufficient for having F, it follows that D_i grounds F. The reason why F can't inherit the causal work of P is simple: *because grounding is asymmetric*. F ontologically depends on the causal work that P does, so the supposition that F inherits this work from P is incoherent. Similar considerations apply *mutatis mutandis* to the difference-making strategy. For familiar reasons, F will make differences to the manifestations of the dispositions in D_i that P doesn't. But from this we can't infer that F does any of the causal work required for those manifestations, since that work consists in grounding D_i, and again, since grounding is asymmetric and D_i grounds F, this is work that F simply can't do.[51]

At this point one might suggest that expecting mental properties to do causal work in my sense is expecting too much.[52] If this is to provide for a reply to the exclusion argument, then *if* we understand a property's having causal powers in terms of its doing causal work, and understand causal work in terms of grounding dispositions, then we'll have to reformulate (AD), because as things stand (AD) requires real, irreducible properties to be irreducible *and to have causal powers*. This amounts to accepting different standards for functional properties and at least some physical properties. Physicists, for instance, commit to properties on the basis of genuinely novel causal work, but perhaps these standards aren't appropriate in psychology. This approach attempts to solve the exclusion problem by admitting that mental properties do no causal work, then arguing that this doesn't result in their elimination because the conditions for ontological commitment to such properties are looser than those for physical properties. Setting psychology aside, science isn't all physics, and there are plenty of functional properties in biology and neuroscience that don't earn their ontological keep in the same way their realizers do, but to whose existence biologists and neuroscientists are nonetheless committed.

I have a certain sympathy with this way of thinking, but it won't do anything to convince the determined exclusionist, who will simply reply that when it gets down to brass tacks, there aren't really any biological or neuroscientific properties either. Retreating to the position that functional properties earn their keep by making a difference, despite the fact that they do no causal work, reduces the disagreement between exclusionists and Humeans, which used to be about whether there's any causal work left for functional properties to do, to one about what our standards for ontological commitment to properties ought to

50. Crane makes a similar point concerning counterfactuals and the causal efficacy of properties in his (2008). Crane's arguments for this conclusion aren't related to mine.

51. It should be noted that my arguments against both the over-determination and difference-making strategies tell only against the use of these strategies to account for the efficacy of functional (*i. e.*, dispositional) properties.

52. LePore and Loewer (1987).

be. Far better for Humeans to grant exclusionists the more stringent condition and then show that functional properties meet it. In the remainder of this paper, I argue that higher-order functional properties can do the same *kind* of causal work — but not, as the arguments of this section show, the same *work* — as their physical realizers.

4. Mental Causation for Functionalists

The causal work of properties consists in synchronically grounding the dispositions of their bearers; finding such work for functional properties to do is harder than showing that their instances are diachronic causes. I'll now briefly recast the exclusion argument in terms of causal work. In Section 2.1, I defined closure (CC) as the claim that every physical effect has a complete, sufficient physical cause, so that I could later define 'complete' in terms of my preferred conception of causal work: put simply, a complete, sufficient physical cause of some event is a sufficient cause of it whose physical properties ground all the relevant dispositions. Think again of a vase broken by the impact of a hammer. Since dispositions manifest reciprocally, we must include in the cause the intrinsic physical properties of the vase itself, as well as those of the hammer, since the former do some of the causal work required if the hammer is to break it. Closure implies that the physical properties of hammer and vase ground both the hammer's disposition to break that kind of vase, and the vase's disposition to be broken by that kind of hammer. A physical cause can be sufficient for some effect without being complete. Suppose emergent downwards causation is possible, whereby a supervenient property-instance emerges from a physical base property-instance, but contributes novel causal powers not grounded by its physical base. Given the sufficiency of the physical base for the emergent property, the effects of emergent powers have *sufficient* physical causes, by transitivity of sufficiency; but they don't have *complete* physical causes.[53]

53. As I understand it, (CC) claims not only that physical effects have sufficient physical causes, but also that the properties that ground such causal relations are physical. So understood, (CC) is about both diachronic causation (causal sufficiency) *and* synchronic causal work (completeness). See Yates (2009) for detailed discussion.

Now if we interpret Alexander's Dictum (AD) as the claim that properties must do at least some causal work (novel or otherwise) in order to earn their ontological keep, it seems to follow straight away that functional properties ought to be eliminated, since: (i) supervenient properties in general can't do causal work their physical bases don't do, since that would render them ontologically emergent, violating causal closure; (ii) functional properties can't do the same causal work as their realizers, since they *are* (at least some of) that work. True, functional properties make a difference their realizers don't, but that doesn't matter, because making a difference isn't sufficient for doing causal work, and it's precisely because they seem to do no causal work that functional properties face the threat of elimination. I now present a theory according to which functional properties do *novel* causal work without grounding any dispositions that aren't grounded in the physical, and so without doing any causal work that physical properties don't do. According to this theory, the causal novelty of functional properties consists not in the dispositions they ground, but in the level, within a hierarchy of grounding relations, from which they ground them. I call it *upwards causation*.

4.1. Upwards Causation[54]

I will say that a mechanism X is individuated by (i) possession of a set D of dispositions, and (ii) a set of components $\{x_1, \ldots, x_n\}$, each having further sets of dispositions $\{D_1, \ldots, D_n\}$, such that $\{x_1, \ldots, x_n\}$ having $\{D_1, \ldots, D_n\}$ constitutively explains why X has D. Suppose I wish to construct a mousetrap using the following items: (a) a 1,000,000V battery, (b) conducting wire of resistance 10Ω, (c) a thin sheet of copper of negligible resistance. I first cut the wire in two, and attach one end of one piece to the positive terminal of the battery, one end of the other to the negative. I then attach the other end of the positive wire to the

54. The account I give here owes much to Craver's (2007), in particular chapters 4 and 5, but I don't attribute what follows to him.

copper sheet, and set the sheet in a housing so that the unattached end of the negative wire is 1cm away from the sheet. The components of my mechanism have the following dispositions: (D_1) the battery is disposed to send a current of (1,000,000/R)A through a conductor of resistance R connected across its terminals; (D_2) the copper sheet is disposed to flex by 1cm when a force of 1N or greater is applied; (D_3) the wire is disposed to conduct a current of (V/10)A when a potential difference of V is applied to it. The mechanism as a whole — my mousetrap — is disposed to shock mice of mass greater than 100g when they sit on the copper sheet. Label this disposition D_M. A mouse of mass 200g sits on the copper sheet, meeting the stimulus condition of the disposition of the sheet to flex by 1cm, bringing it into contact with the wire. This in turn stimulates the reciprocal dispositions of the battery and wire, sending a current of 100,000A through the sheet, which shocks the mouse. The importance of structure should be clear. Were the negative wire not located 1cm below the copper sheet, the sheet's manifesting its disposition to flex by 1cm when an object of mass greater than 100g rests on it would not trigger the dispositions of the other components. The way the components are structured enables the manifestation of one disposition to be the stimulus of another. Of course not every mechanism is structured in such a linear way, but I needn't consider more complex cases to make my point. The entity that has D_M is the mousetrap itself: D_M isn't identical to any disposition that can be attributed to the individual components, even though its manifestation on some occasion is constituted by the structured manifestations of the dispositions of the components.

Now consider: which properties do the causal work of grounding D_M? A key element of my proposal is that D_M is grounded in the dispositions of the mousetrap's components, together with their spatiotemporal structure. According to (DISP), the battery's having D_1 consists in there being a set of possible background conditions $\{B_1, \ldots, B_n\}$, such that if the battery were in any of the B_i and a conductor of resistance R were connected to its terminals, it would conduct a current of (1,000,000/R)A. The copper sheet's having D_2 consists in there being a set of possible conditions $\{B_1', \ldots, B_n'\}$ such that were the sheet in any of the B_i' and a force of 1N were applied, it would flex by 1cm. If the intersection of these two sets is empty, the mechanism won't work. When designing a mechanism, we maximise the intersection set and make sure it corresponds to the conditions under which we want the mechanism to work. This is why it matters not only *that*, but also *how*, the dispositions of the components are realized. My mousetrap has D_M because: (1) the components have D_1–D_3; (2) the components have a certain spatiotemporal structure; (3) D_1–D_3 have a non-empty intersection set of background conditions. Call (1)–(3) the *dispositional structure* of the mousetrap. Such structures typically involve components which are themselves mechanisms, and have their dispositional properties in virtue of their own dispositional structures. As in the example of the hammer and vase in Section 3.2, the properties that ground D_M thereby ground potential causal relations. The mousetrap's having D_M consists in its having intrinsic properties — its dispositional structure — in virtue of which there's a range of possible conditions under which, were a mouse to sit on the copper sheet, it would be shocked. What grounds D_M thereby grounds potential difference-making causal relations: whether or not the mouse steps on the sheet makes a difference to whether or not it's shocked.

An obvious rejoinder: Why isn't it the physical properties of its components that do the causal work of grounding the dispositions of a mechanism? Reply: *It is!* The causal powers of a mechanism are grounded in the fundamental physical properties of its components. However, I deny that this leaves no causal work for functional properties to do. In fact, I claim that the grounding of a mechanism's dispositions by the basic physical properties of its components is secured only by the transitivity of grounding, and the fact that such dispositions are grounded in the dispositional properties of its components. Think again of the mousetrap, and focus on the copper sheet, and the physical properties of the sheet in virtue of which it has D_2. Suppose the copper sheet breaks and I need to replace it but have only a thick, inflexible

aluminium sheet to hand. What I must do, if I want my mousetrap to continue working, is to file the sheet to a thickness such that it too flexes by 1cm when a force of 1N is applied. Granting for argument's sake that a suitably thin sheet of aluminium also has negligible resistance, this example shows that what's important when it comes to grounding D_M is that the components are appropriately structured and have D_1–D_3 (realized so that their background conditions overlap). Holding the mousetrap's other components fixed, and removing the copper sheet, what we must do to get it working again is install a component of negligible resistance with D_2. Put differently: in virtue of its intrinsic nature, the copper sheet has D_2, and in virtue of having D_2 it's capable, when placed in appropriate structural relations with the other components, of completing the mechanism. The causal work that the physical properties of our replacement sheet must do in order to partially ground D_M is to ground D_2, the having of which makes it possible for the sheet to form part of a dispositional structure that constitutively explains why the mousetrap has D_M.

4.2. Proximal and distal grounding

The physical properties of components in a mechanism do the causal work of grounding the dispositions of those components, which then ground those of the mechanism. The same is true even at the level of fundamental physics. An electron orbits a proton in the hydrogen atom due to the dispositional structure of electron and proton. The physical properties of electrons and protons ground the dispositions of the hydrogen atom, by grounding the reciprocal dispositions of electron and proton to attract each other. It's unlikely that the dispositions grounded by fundamental properties such as charge are multiply realizable. It remains true, however, that fundamental particles combine into more complex mechanisms *in virtue of dispositions* that they have in virtue of their fundamental properties. Assuming the notion of grounding to be sufficiently well understood, we can distinguish two kinds of grounding: *proximal* and *distal*. Where P and P′ range over properties (including dispositional structures), *X* over particulars (including mechanisms), and D over dispositions, we may define proximal grounding as follows:

(PG) *X*'s having P proximally grounds *X*'s having D iff (i) *X*'s having P grounds *X*'s having D, (ii) there is no P′ such that: (a) *X* has P′, and (b) *X*'s having P grounds *X*'s having P′, and (c) *X*'s having P′ grounds *X*'s having D.

Intuitively, proximal grounding is grounding without intermediaries. My mousetrap has many levels of mechanism: the battery, for instance, is a mechanism, which is disposed to send a current of (1,000,000/R)A through a conductor of resistance R connected to its terminals, in virtue of *its* dispositional structure — being composed of an anode and cathode, separated by an electrolyte, say. The same goes for the copper sheet, which is disposed to flex by 1cm when a force of 1N is applied in virtue of the way its atoms are arranged, their dispositions to exert certain forces on their neighbours, and so forth. My mousetrap has dispositional structures at lower levels of mechanism than the one I've been discussing so far, which don't proximally, but distally, ground its power to shock mice:

(DG) *X*'s having P distally grounds *X*'s having D iff (i) *X*'s having P grounds *X*'s having D, and (ii) ¬(*X*'s having P proximally grounds *X*'s having D).

Grounding *simpliciter*, construed as either proximally or distally grounding a disposition, is transitive. Proximal grounding is intransitive, because by definition if *a* proximally grounds *b* and *b* proximally grounds *c*, then *a* distally grounds *c*; distal grounding is transitive. The basic physical properties of a mechanism at best *distally* ground its characteristic dispositions. The role of fundamental physical properties is to proximally ground the powers of fundamental physical particles. Such particles combine into mechanisms (atoms) which have certain characteristic powers in virtue of their dispositional structures. These in turn combine into molecules, chemical compounds, cells, and so on all the way up. The fundamental properties of my mousetrap distally

ground its power to shock mice, which leaves plenty of causal work left for the functional properties of its components to do — *proximally* grounding this power. Proximal and distal grounds occupy different places in a hierarchy of grounding relations through which the causal influence of fundamental physics extends upwards to medium-sized dry goods. This doesn't involve causal closure violations, because there's no pressure at all to read the notion of causal work implicit in (CC) in terms of proximal grounding. Indeed, read in this way, (CC) is false, for there are many dispositional structures between basic physics and the powers of ordinary physical particulars. Conversely, if we take (CC) to imply that the powers of ordinary things — agents, engines, aeroplanes, batteries, bananas — to bring about certain physical effects are grounded in their fundamental physical properties, we must read 'grounded' as "distally grounded". Far from *precluding* the causal novelty of functional properties, (CC) actually *requires* it. Functional properties at a specific level of mechanism are causally novel because they occupy a unique place in hierarchy of dispositional structures, without which basic physical properties could ground only the powers of basic physical particulars.

Upwards causation may initially strike the reader as similar to the kind of over-determination strategy I rejected in Section 3.4. There, I argued that it's incoherent to suppose that functional properties inherit the causal work their realizers do, because they are at least part of that work. Since I hold that functional properties proximally ground dispositions that are distally grounded in the physical, however, I must accept that some causal work is, in a sense, over-determined. For upwards causation to work, there must be dispositions that are grounded *simpliciter* (but not proximally) in both functional properties and their basic physical realizers, so grounding *simpliciter* is sometimes over-determined. This doesn't undermine my previous arguments against over-determination, because it remains the case if functional properties simply inherited the causal work of their realizers, they would be self-grounding dispositions, which doesn't make sense. There's nothing wrong with supposing that there are causal powers grounded in both functional properties and their realizers; what's incoherent is the stronger claim that the former inherit the causal work of the latter. Indeed, given upwards causation, when basic physical properties *distally* ground a power via functional intermediaries, it makes more sense to say that the physical properties inherit this component of their causal work from the functional properties that *proximally* ground it.

4.3. Upwards Causation and the Mental

In order to save functionalist mental properties from the threat of elimination, we need to find novel causal work for them to do. They can't do the same causal work as their realizers, because they *are* the causal work their realizers do. We are now in a position to find such work for functional mental properties to do.[55] Mental properties proximally ground *agent-level dispositions* not proximally grounded in agents' physical properties. They are distally so grounded, but this depends on the transitivity of grounding and the fact that such dispositions are proximally grounded in the mental. Suppose for the sake of argument that the ambient temperature comes in two flavours: cold and warm. I have the following mental properties: (i) I want to be neither too hot nor too cold, (ii) I want to go outside, (iii) I believe that if it's cold outside, then if I go outside I'll be too cold unless I dress warmly, (iv) I believe that if it's warm outside, then if I go outside I'll be too hot unless I don't dress warmly. Given functionalism, (i)–(iv) are (clusters of) dispositions. For instance, (i) is — *inter alia* — the disposition to dress warmly if I have states (ii) and (iii) and believe it's cold outside. What I don't yet know is whether it's cold or warm: I've only just got up and haven't yet had my coffee, which comes before opening the door to check the temperature. Being a comparatively normal, rational agent with properly functioning senses, I'm disposed to believe that it's cold outside when it *is* cold; ditto warm. These states combine to ground the following agent-level dispositions: (a) I'm disposed to dress warmly

55. What follows isn't intended as an exhaustive account of the causal novelty of such properties.

when it's cold outside, (b) I'm disposed not to dress warmly when it's warm outside. Label these D_C and D_W respectively.

My having D_C consists in my having intrinsic properties in virtue of which there's a range of nomically possible conditions in which I would dress warmly if it were cold outside; similarly *mutatis mutandis* for D_W. These conditions, as before, depend on those of the grounding dispositions — including, for example, that I'm not under the influence of drugs such that if it were cold outside, I wouldn't believe it. The intrinsic properties that ground D_C and D_W thereby ground potential causal relations. Suppose it's cold, that some relevant background conditions for D_C and D_W obtain, and I dress warmly. The fact that it's cold is a difference-making cause of my so dressing. The nearest worlds at which it's cold will be worlds at which I have D_C and relevant background conditions obtains, hence worlds at which I dress warmly. The nearest worlds at which it's not cold will be worlds at which I have D_W and relevant background conditions obtain, hence worlds at which I don't dress warmly. The psychological dispositional structure that grounds D_C and D_W thereby grounds a range of potential difference-making causal relations between the ambient temperature and my attire. This is no philosopher's invention: I really do have D_C and D_W, in virtue of something like the dispositional structure outlined, and my attire really is counterfactually correlated with the ambient temperature.[56] Psychology proximally grounds causal relations which, since physics grounds psychology, are distally grounded in physics. Since, by anyone's lights, a novel causal role is sufficient for robust ontological commitment, functional properties are as non-redundant as the properties of fundamental physics. Mechanisms at any level have their defining dispositions proximally grounded in the dispositions of their components; that such mechanisms are ultimately grounded in fundamental physics is only possible at all because of the way in which dispositions compose. Basic physical properties ground the powers of complex mechanisms by grounding those of their most basic components; from there on it's functional all the way up.

The theory detailed above involves treating mental properties as properties of components in psychological mechanisms. It might be objected that mental properties are properties of agents, and are therefore at the same level of mechanism as the agent-level dispositions I claim they ground. A first rejoinder: being properties of agents doesn't preclude mental properties from grounding further dispositions at the agential level. I see no reason why there shouldn't be dispositional structures in which a single component has a set of dispositions which constitutively explain why that very component has some further disposition. Still, that isn't how I see functionalist mental properties, nor, arguably, is it what functionalists ought to say. Functional properties are physically grounded dispositions. If the physical realizers of functional properties are neural properties, then it seems the brain states that bear them will also bear any dispositions they ground. On my account, functional property F_i is the property of being an x such that there are properties $P_1, \ldots, P_n$ in virtue of which, collectively, a certain conjunction of subjunctives is true, and x has P_i. Treating functional realization as a kind of same-subject necessitation is commonplace, and one which entails that if P_i is a brain property, so is F_i. But on reflection, why shouldn't a property of my brain also be a property of *me*? If I grow my fingernails long, isn't it true both that my fingers have long nails, and that I do? One might wish to insist that mental properties are properties of whole agents, and not of any proper parts thereof, but in that case functionalists must either say the same about their realizers, or else find an alternative to the standard quantificational way of defining second-order properties. What functionalists ought to say, I submit, is that agents have mental properties in the same derivative way that agents have fingernails. Perhaps there's some reason for thinking

56. Note that this structure also grounds psychological difference-making causes. The ambient temperature makes a difference to the way I dress because I have reliable *belief*-forming mechanisms that enable me to detect it. As well as D_W, for instance, I also have — in virtue of the dispositional structure given by (i)–(iv) — the disposition to dress warmly if I *believe* that it's cold outside. My belief that it's cold outside now will later be a difference-making cause of my dressing warmly before leaving work, as will the various other dispositions in the structure.

that mental properties can't both be properties of agents and their brains, but I don't see what it could be.

4.4. Causal exclusion bites back?

I claim that the causal work dispositions do consists in their grounding dispositions, and thereby causal relations, at higher levels of mechanism. But thus far I've said nothing to explain why we should believe there are such dispositions to ground. Given physicalism, it follows that every mechanism has a fundamental dispositional structure, consisting in its being composed of fundamental physical components having certain dispositions, and a certain spatiotemporal structure. The basic components of each mechanism will have their dispositions in virtue of their fundamental physical properties, and these dispositions, together with the way their bearers are structured, are sufficient to fully explain everything the mechanism does. Suppose a proton and an electron combine, in virtue of the dispositions grounded in their respective charges, to form a hydrogen atom. Such an atom is a mechanism, in my sense, whose dispositional structure — its being composed of a suitably disposed and structured electron and proton pair — grounds dispositions such as its being combustible. Suppose a sample of hydrogen combusts on some occasion, under circumstances C. The dispositions of the electron and proton, together with their spatiotemporal relations, are sufficient to *explain* the sample's combustion in C. The same will be true all the way up. Given any mechanism, however complex, we will in principle be able to explain what it does on some occasion in terms of the manifestation of the dispositions of its fundamental physical components, and their spatiotemporal relations.

The explanatory adequacy of fundamental dispositions licenses the following objection: (A) we shouldn't posit any dispositions we don't need in order to explain why things behave the way they do, and (B) we'll never need to posit higher-order dispositions for explanatory purposes. If this objection is correct, then fundamental physical properties are the only ones that do causal work, because the dispositions of fundamental physical particulars are the only causal powers to ground.[57] Clearly, it's no use arguing that dispositions do novel causal work in proximally grounding higher-order dispositions if there aren't any higher-order dispositions to ground. It seems that in order to show that there's causal work for any dispositions to do, I need to assume the reality of higher-order dispositions, but that's exactly what's at issue.

It's important to get clear about the dialectic before proceeding. On the table is the claim that any event we can explain in terms of the manifestation of a higher-order disposition can be explained in terms of the structured manifestations of fundamental physical dispositions. I assume the reality of agent-level dispositions, and argue that functionalist mental properties are causally novel in virtue of proximally grounding such dispositions. If mental properties are causally novel, there's no question of their being eliminated, so my strategy aims to secure mental properties via novelty of *synchronic causal work*. But the objector doubts the reality of agent-level dispositions, and so of course won't grant me that mental properties ground them. Non-fundamental dispositions lack *diachronic causal-explanatory novelty*, and so should themselves be eliminated. Without such dispositions, there's nothing for functional properties to ground, and therefore no upwards causation. Fundamental dispositions secure their ontological status, if at all, not by doing the causal work of grounding further dispositions, but by dint of their diachronic causal-explanatory role.

I reply that Humeans have already shown how to rebut this objection: whether or not a vase is fragile makes a difference to whether or not it breaks, but whether or not it has *this* basic dispositional structure doesn't. If any kind of causal explanation is contrastive, then higher-order dispositions aren't explanatorily redundant, because they have a contrastive explanatory role that fundamental dispositions don't. If we

57. Something like this underpins Merricks' (2003) eliminativism about ordinary objects. Any irreducible causal powers we might attribute to such objects are rendered otiose by the causal powers of their ultimate constituents. Lacking novel causal powers, the objects themselves should be eliminated.

want to explain why the vase broke rather than not, it's no use citing its basic dispositional structure if we also think that at the closest possible worlds where it lacks that specific structure, it still breaks. The vase's basic dispositional structure will be required to explain the precise *manner* of its breaking, but unless contrastive causal explanation is itself dispensable, higher-order dispositions like fragility aren't diachronically redundant. This is of course the "dual explanandum strategy", which can also be employed as a direct response to the causal exclusion problem.[58] I don't employ the dual explanandum strategy in this way, because Humean causation isn't sufficient for causal work, and the exclusion problem is precisely the problem of finding such work for functional properties to do. My aim here is to (i) allow for the sake of argument that if higher-order dispositions were diachronically redundant, then we'd have grounds for their elimination, and (ii) employ the dual explanandum strategy to show that they aren't redundant in this sense. Having a novel difference-making role isn't sufficient for doing causal work (novel or otherwise), but it *is* sufficient for the kind of causal-explanatory relevance we need to block explanatory redundancy arguments. Similar arguments can be run to show that macro-properties in general — whether dispositional or not — aren't explanatorily redundant: *striking* a vase makes a difference to its shattering that a particular *microphysical realization* of striking doesn't make.

Exclusion, however, still isn't done biting. All I've done so far is block a redundancy argument to the effect that higher-order dispositions aren't real. If such properties aren't real, I claim, it isn't because they lack diachronic explanatory novelty. It's another matter, however, to show that they *are* real, and it remains the case that my theory presupposes them. Worse than that, I need macro-events such as its being cold, dressing warmly, etc. as the relata of the difference-making causal relations that license the claim that grounding higher-order dispositions counts as causal work in the first place. This objection gets the burden of proof all wrong: the functional properties the exclusion argument targets *are* higher-order dispositions. Those who run Kim's exclusion argument against functional properties can't *assume* that functional properties don't exist, because that's what the argument is supposed to *show*. And since functional properties are defined in terms of macro-events, the exclusion argument can't be premised on their non-existence either.

We can think of the exclusion argument as a *reductio*. First, assume that there are functional properties. Then show that such properties can't do causal work, and conclude that they don't exist after all, because if they did, there would be causally redundant properties, which (AD) rules out. The argument therefore depends on the truth of the following subjunctive: *if functional properties existed, there wouldn't be any causal work for them to do*. And that's exactly what upwards causation refutes: if there are functional properties, there's plenty of causal work left for them to do, viz., proximally grounding functional properties at the next level up. If I were arguing *that functional properties exist*, this would be bootstrapping, but I'm not, so it isn't. The diachronic explanatory novelty of functional properties is sufficient to justify our belief that they exist, so the burden of proof rests with those who would argue that they don't. Without tacitly assuming that there are no functional properties, proponents of the causal exclusion argument can't show that functional properties have no causal work to do. But if they are prepared to assume that functional properties aren't real, what's the exclusion argument for? Perhaps the reader has a nagging suspicion that mental properties still aren't doing any *essential* causal work. Couldn't the agent-level dispositions of agents be grounded solely by properties of basic physics? What if one agrees with me that agents have agent-level dispositions, but denies the reality of mental properties? Well, mental properties *are* dispositions (of components in psychological mechanisms, which contribute to grounding the dispositions of agents). There's no obvious reason to allow that there are agent-level functional properties, such as the disposition to dress warmly when it's cold outside, but deny that there

58. Marras (1998).

are functional mental properties. Whatever the current objector's reason for doubting the existence of mental properties, it had better not be their dispositional nature. If I'm right that mental properties form part of a hierarchy of dispositional structures through which basic physical properties ground the powers of agents, then their grounding roles are every bit as important as those of their ultimate physical grounds.

As I see it, there are as many levels of dispositional properties as there are levels of mechanism. Protons and electrons are disposed to form Hydrogen, in virtue of their basic physical properties. Hydrogen has the further disposition to combust under certain circumstances, releasing water and energy. Now suppose we make a combustion engine in which Hydrogen is a component — the *fuel*. The engine's having the power to make the vehicle move will depend *inter alia* on the dispositional properties of Hydrogen, but that's just to say that the basic physical properties of the electrons and protons that compose the Hydrogen in the mechanism *distally* ground the power of the engine to make the vehicle move. Causal closure entails that all dispositions are grounded in properties of basic physics, but not that there are no intermediate dispositions. Mental properties, I claim, are among those intermediates, and so are as important to the workings of the agent as the combustibility of hydrogen is to an engine that burns it as fuel.

4.5. Conclusion

We live in a world where all causal powers, hence all causal relations, are grounded in fundamental physics, and therefore dependent upon it. This is apt to make it seem as though there isn't really anything but physics. If the physical is doing all the causal work, why bother with anything else? The central contention of this paper is that if there are higher-order causal powers to ground, then although such powers are distally grounded in the physical, distal grounding isn't all the causal work there is. Agent-level dispositions, for instance, are distally grounded in the physical through a hierarchy of nested mechanisms, culminating in the psychological mechanisms whose functional properties proximally ground those dispositions. Fodor says: [59]

> So, then, *why is there anything except physics*? That, I think, is what is *really* bugging Kim. Well, I admit that I don't know why. I don't even know how to *think about* why. I expect to figure out why there is anything except physics the day before I figure out why there is anything at all, another (and, presumably, related) metaphysical conundrum that I find perplexing.

I agree with Fodor that this is what's bugging Kim — and Merricks, and Heil[60] — and while I share Fodor's pessimism about the prospects for an answer, I see no compelling argument for the conclusion that there *isn't* anything but physics. Extant Humean responses to the exclusion problem are, to register my agreement with Kim, a free lunch:[61] making a difference isn't sufficient for doing the kind of causal work that causal closure appears to render the province of fundamental physics alone, and the thought that functional properties do the same causal work as their realizers is incoherent. However, the distinction between proximal and distal grounding, together with the reality of higher-order mechanisms and their characteristic dispositions, enables functional properties to pay for their lunch the same way physical properties do. The causal structure of the world, in my view, is irreducibly layered. There's no causal exclusion problem because there's far more causal work to do, in constructing such a world, than is commonly supposed.[62]

59. Fodor (1997) p. 161.

60. Merricks (2003); Heil (2003).

61. Kim (1998), ch. 3.

62. This paper grew out of a seminar at King's College London, in which I tried to persuade Jim Hopkins of the intractability of the exclusion problem. My thanks to Jim for lively and thought-provoking opposition. Thanks also to Mahrad Almotahari, Phillip Goff, Chris Hughes, Nick Jones, Shalom Lappin, Mark Textor, Raphael Woolf, and two anonymous referees. Based on research funded by a British Academy Postdoctoral Fellowship.

Bibliography

Bennett, K. (2003). "Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract It", *Noûs* 37, pp. 471–497.

Bird, A. (1998). "Dispositions and Antidotes", *Philosophical Quarterly* 48, pp. 227–234.

—— (2007). *Nature's Metaphysics*. Oxford University Press.

Block, N. (2003). "Do Causal Powers Drain Away?", *Philosophy and Phenomenological Research* 67, pp. 133–150.

Crane, T. (2008). "Causation and Determinable Properties: On the Efficacy of Colour, Shape and Size", in Hohwy and Kallestrup (2008), pp. 176–195.

Craver, C. (2007). *Explaining the Brain*. Oxford University Press.

Fara, M. (2005). "Dispositions and Habituals", *Noûs* 39, pp. 43–82.

Fodor, J. (1997). "Special Sciences: Still Autonomous After All These Years", *Noûs* 31, Supplement: *Philosophical Perspectives 11, Mind, Causation and World*, pp. 149–163.

Hawthorne, J., and D. Manley (2005). "Mumford's *Dispositions*", *Noûs* 39, pp. 179–105.

Heil, J. (2003). *From an Ontological Point of View*. Oxford University Press.

—— (2005). "Dispositions", *Synthese* 144, pp. 343–356.

Hohwy, J., and J. Kallestrup, eds. (2008). *Being Reduced: New Essays on Reduction, Explanation and Causation*. Oxford University Press.

Jackson, F. (1996). "Mental Causation", *Mind* 105, pp. 377–413.

Johnston, M. (1992). "How to Speak of the Colours", *Philosophical Studies* 68, pp. 221–263.

Kallestrup, J. (2006). "The Causal Exclusion Argument", *Philosophical Studies* 131, pp. 459–485.

Kim, J. (1992a). "The Nonreductivist's Troubles with Mental Causation", in J. Heil and A. Mele, A. (eds.), *Mental Causation*. Oxford University Press. Reprinted in Kim (1993). *Supervenience and Mind*. Cambridge University Press, pp. 336–357.

—— (1992b). "'Downward Causation' in Emergence and Nonreductive Physicalism", in A. Beckermann, H. Flohr, and J. Kim (eds.) (1992). *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. Berlin: De Gruyter.

—— (1998). *Mind in a Physical World*. MIT Press.

—— (2003). "Blocking Causal Drainage and Other Maintenance Chores with Mental Causation", *Philosophy and Phenomenological Research* 67, pp. 151–176.

—— (2005). *Physicalism, Or Something Near Enough*. Princeton University Press.

LePore, E., and B. Loewer (1987). "Mind Matters", *Journal of Philosophy* 84, pp. 630–642.

Lewis, D. (1986). "Events", in *Philosophical Papers Vol. II*. Oxford University Press, pp. 241–269.

—— (1997). "Finkish Dispositions", *Philosophical Quarterly* 47, pp. 143–158.

List, C., and P. Menzies. (2009). "Non-Reductive Physicalism and the Limits of the Exclusion Principle", *Journal of Philosophy* 106, pp. 475–502.

Marras, A. (1998). "Kim's Principle of Explanatory Exclusion", *Australasian Journal of Philosophy* 76, pp. 439–451.

Martin, C. (1994). "Dispositions and Conditionals", *Philosophical Quarterly* 44, pp. 1–8.

—— (2007). *The Mind in Nature*. Oxford University Press.

McKitrick, J. (2005). "Are Dispositions Causally Relevant?", *Synthese* 144, pp. 357–371.

McLaughlin, B. (2006). "Is Role-Functionalism Committed to Epiphenomenalism?", *Journal of Consciousness Studies* 13, pp. 39–66.

Menzies, P. (2008). "The Exclusion Problem, the Determination Relation, and Contrastive Causation", in Hohwy and Kallestrup (2008), pp. 196–217.

Merricks, T. (2003). *Objects and Persons*. Oxford University Press.

Molnar, G. (2003). *Powers: A Study in Metaphysics*. Oxford University Press.

Mumford, S. (1998). *Dispositions*. Oxford University Press.

Papineau, D. (2010). "Realism, Ramsey Sentences and the Pessimistic Meta-induction", *Studies in History and Philosophy of Science* 41, pp. 375–85.

Prior, E., R. Pargetter, and F. Jackson (1982). "Three Theses About Dispositions", *American Philosophical Quarterly* 19, pp. 251–257.

Segal, G., and E. Sober (1992). "The Causal Efficacy of Content", *Philosophical Studies* 63, pp.1–30.

Shoemaker, S. (1980). "Causality and Properties", in P. van Inwagen (ed.), *Time and Cause*. Dordrecht: Reidel.

—— (2001). "Realization and Mental Causation", in C. Gillett and B. Loewer (eds.), *Physicalism and its Discontents* (Cambridge: Cambridge University Press), pp. 74–98

Sider, T. (2003). "What's So Bad About Over-determination?" *Philosophy and Phenomenological Research* 67, pp. 719–726.

Stephan, A. (1997). "Armchair Argument Against Emergentism", *Erkenntnis* 46, pp. 305–314.

Wilson, J. (2002). "Causal Powers, Forces and Superdupervenience", *Grazer Philosophische Studien* 63, pp. 53–78.

Witmer, G. (2003). "Functionalism and Causal Exclusion", *Pacific Philosophical Quarterly* 84, pp. 198–214.

Yablo, S. (1992). "Mental Causation", *Philosophical Review* 101, pp. 245–280.

Yates, D. (2009). "Emergence, Downwards Causation and the Completeness of Physics", *Philosophical Quarterly* 59, pp. 110–131.