# Mechanisms, Causes, and the Layered Model of the World*

**Stuart Glennan**
**Department of Philosophy & Religion**
**Butler University**

Forthcoming in *Philosophy and Phenomenological Research*

**Abstract**

Most philosophical accounts of causation take causal relations to obtain between individuals and events in virtue of nomological relations between properties of these individuals and events. Such views fail to take into account the consequences of the fact that in general the properties of individuals and events will depend upon mechanisms that realize those properties. In this paper I attempt to rectify this failure, and in so doing to provide an account of the causal relevance of higher-level properties. I do this by critiquing one prominent model of higher-level properties – Kim's functional model of reduction – and contrasting it with a mechanistic approach to higher-level properties and causation.

* This paper was originally delivered to a symposium on Mechanisms in the Sciences at the APA Central Division Meetings in Spring 2006. I am grateful to the audience and to my cosymposiasts Peter Machamer, Stathis Psillos and Bill Wimsatt for their input. I would especially like to thank Carl Craver for his commentary at that session and for the many discussions we have had about the topic of this paper. The final version also benefited from helpful comments by an anonymous referee.

The distinction between an individual and its properties, a substance and its accidents, particulars and universals, or whatever you will call them, is very deeply entrenched in our philosophical worldview. This is a consequence of our philosophical history – going back most prominently to Aristotle. More broadly it is a consequence of the structure of our language and our logic. The distinction between concept and object lies at the center of Frege’s semantics, where logic slides into metaphysics.

The distinction between individuals and their properties plays a prominent role in the history of the debate over the nature of causal relations. In Hume’s analysis for instance, causation is a relation between objects that are sorted by properties. Most contemporary accounts of causation suppose that individuals and the events they participate in stand in causal relations in virtue of laws of nature, where laws of nature are characterized in terms of relations between properties or in terms of generalizations about individuals that are the bearers of properties.

Notwithstanding its ancient pedigree and continuing influence, this ontological framework does not sit well with another pervasive ontological assumption – what we might broadly call the layered model of the world. This view, which has its origins in ancient atomism and which blossomed into the mechanical philosophy of the seventeenth century, holds that the entities which we see in the world are typically composed of smaller entities, with these entities being composed in turn of smaller entities, and so on until one reaches some bottom level of entities – the “atoms.” This view moreover holds that the macro-world, including its entities, and their properties and causal relations, is in some sense dependent on and explicable in terms of the micro-world. The layered model does not mesh with the object/property/law ontology because that ontology is essentially flat and seems to suppose that what the objects, properties and laws are are the basic facts about the universe. This disconnect becomes especially problematic when we seek to understand the nature of causal relations.

My aim in this paper is to square the object/property/law ontology with the layered model by construing the layers that make up the world in terms of nested mechanisms, and in so doing to help to resolve concerns that higher-level causes are epiphenomenal. I will argue that much of the literature on higher-level causation suffers from what I'll call *property bias* – a tendency to think about the properties of objects as basic facts rather than mechanically explicable facts and a tendency to think about causal relationships as holding directly between properties rather than being mediated by particular mechanisms.

My paper is in four parts. In the first, I offer a simple analysis of the character of causal claims. I suggest that causal claims are of two kinds – claims about the objects and events that *produce* effects and claims about the properties of or facts about these objects and events that are *relevant* to these effects. In the second part I contrast the traditional law-based approach to thinking about causally relevant properties with one in which laws and properties are explained by mechanisms, arguing that law-based accounts which fail to take seriously the layered model yield flawed accounts of causal relevance. In the third part I offer criticisms of one of the most serious attempts to integrate a property-based view of causal relations with the layered model – Jaegwon Kim's functional model of reduction. In the final part I spell out an alternative to the functional model of reduction – one that avoids property bias and thereby provides a more satisfactory account of higher level causation.

# I. Productivity and Relevance

No account of the relation of higher to lower level causes will get very far without some preliminary decisions on what kind of things causes should be taken to be. Is causation a relation between objects, events, tropes, aspects, variables, or facts? And what, by the way, are each of these things? That account must, moreover, suggest something about the relationship between singular or token causal claims, and causal claims about types. It is not possible here to enter into an extended discussion about the nature of the causal relation and its relata, so I will accept what I take to be the most commonly held view – that causation is in the first instance a relation between events, and that events are understood in terms of instances or occurrences or exemplifications of properties.[1]

What I do offer here is a proposal for a canonical form of a causal claim. While I make no argument for universality, a wide variety of causal claims can be represented in this form. The particular virtue of this form is that it brings to light an important distinction between two sorts of causes – causally productive events and causally relevant properties. The form is as follows:

[1] (Bennett 1988) provides a useful study of various theories about the semantics and metaphysics of events, and of the relation between events, facts and causes. A briefer and more up-to-date discussion can be found in (Schaffer 2003). While some metaphysicians would find my language here hopelessly sloppy, my major concern in characterizing events is to suggest that events have something to do with properties. I will not worry here about the individuation of events, but I will assume that events must be more finely individuated than space-time locations and more coarsely individuated than facts. If Bob coughs and itches at the same place and time, the coughing and the itching are different events, but if Bob coughs loudly, that is a fact about the coughing event, and not another event.

(C) Event *c* causes event *e* [in background conditions *B*] in virtue of properties *P* [of *c*, *e* or *B*].

For example,

Bob's coughing (*c*) caused Carol to wake up (*e*) in virtue of cough's loudness (*P*).

We often speak of objects rather than events as causes – for instance if someone were to claim that the bomb caused Arnold's death. But this language is easily translated into event language, since, when one speaks of an object causing an event, it is generally the case that an event involving the object caused the subsequent event. For instance, in this case, Arnold's death was caused by the event of the bomb's exploding.

Although the canonical form (C) singles out one cause of an effect, the set of events causally sufficient to bring about an effect are typically large, so that when we speak of *the* cause of an event, we are using pragmatic criteria to single out a certain event as especially salient. When we say that the straw that broke the camel's back caused the camel's back to break, we know of course that the causal responsibility for the back-breaking was shared with whatever else was on the camel's back.

The "in virtue of" clause of the canonical form highlights the importance of *causally relevant properties* to explaining *why* a cause produces an effect. To illustrate the importance of the "in virtue of" clause, consider the claim that Bob's coughing caused Carol's waking in virtue of the cough's loudness. Bob's coughing is an event, and that event produces Carol's waking. But if we know only this much, we are missing important information about the causes of Carol's waking. In particular, we don't know what properties of Bob's cough made Carol wake. If we know that Carol woke because she heard Bob's cough, then it was the loudness of the cough that mattered. But other things about the cough could have mattered. For instance, Carol might have been woken by the cough because some nasty stuff landed on her shoulder.

Causal relevance is essentially a counterfactual notion. To say that the cough's loudness was relevant to Carol's waking is to say that if the loudness had been somewhat different, Carol wouldn't have woken. Similarly, to say that the amount of stuff coughed up wasn't relevant is to say that even if the amount of stuff coughed up had been different, Carol still would have woken.

In this example, we have considered only causally relevant properties of the productive event. Relevance claims can, however, take a much wider variety of forms, and (C) is meant to allow some of this variety to be captured. In the canonical form, we say that *c* causes in virtue of properties either of *c*, of *e* or of background conditions *B.* For instance, not only might we say that Carol's waking is caused by the loudness of Bob's cough, but also by the sensitivity of Carol's ears. Were Carol a bit deafer, she wouldn't have been awoken by Bob's coughing. As an example of a case in which a causally productive relation between events holds in virtue of background conditions, consider the claim that the campfire caused the forest fire in virtue of the low humidity and high winds. The low humidity and high winds are properties neither of the campfire nor the resulting forest fire, but they are features of the background conditions that are causally relevant, in the sense that, had they been different, then the campfire would not have produced the forest fire.

In contrast to claims about causal relevant properties, claims about productive events are not essentially counterfactual. To say that one event produced another is to say that *in fact* the causative event is connected to the effect via a continuous chain of causal processes. It matters not what might have happened had the productive event not occurred. In cases of overdetermination, for instance, it might be the case that had the productive cause not occurred, the effect still would have occurred, but this lack of counterfactual dependence would not undermine the productive relation in the actual case.

A further reason to think that productivity and relevance are distinctive notions is that it is possible to have knowledge of either kind of cause in the absence of the other. Suppose that we discover an outbreak of diarrhea and vomiting occurs within a population of people. We may, using some variant of Mill's methods, quickly surmise that the cause of the illness in each of these cases was the ingestion of food from a particular source. At this point we have information about the productive cause of the illness. We do not however understand which properties of the food led to the illness. Was it a strain of bacteria? An inorganic contaminant?

Conversely, in some cases we may know about properties that are causally relevant without knowledge of the particular productive events. We have very good reason to believe that smoking is probabilistically relevant to lung cancer, and when a smoker gets lung cancer, we have good reason to believe that the smoking was relevant to her contracting cancer. At the same time though, we cannot identify the particular events that got the cancer started. Smoking is not a particular event, and we have no way of telling which cigarettes or even what quantity of cigarettes might have triggered the growth of a particular cancer.[2]

[2] While the particular way I have drawn the distinction between productivity and relevance is somewhat different from other accounts, both concepts have antecedents in the literature, as does the claim that there are two kinds of causes. Various philosophers have argued for the centrality of productive continuity (Bogen 2004; Dowe 2000; Machamer, Darden, and Craver 2000; Salmon 1984). Others have argued for

While these remarks are not sufficient to fully characterize the concepts of productivity and relevance, I hope that they at least suggest that we make two different kinds of causal claims. In the following section, I shall try to spell out the concepts a bit more by comparing two accounts of how we should understand the relationship between causally relevant properties and productive events.

## II. Relevance, Laws and Mechanisms

The canonical form (C) characterizes causation a relation between events that obtains in virtue of certain properties of those events.[3] Why the invocation of properties here? The standard answer is that for one event to cause another event those events must fall under a causal law. This is how Fodor puts it:

> Singular causal statements need to be covered by causal laws. That means something like:
>
> *Covering Principle*: If an event $e_1$ causes an event $e_2$ then there are properties *F*, *G* such that (a) $e_1$ instantiates *F*, (b) e2 instantiates *G* and (c) "*F* instantiations are sufficient for *G* instantiations" is a causal law.[4]

Fodor is hardly alone in thinking this. He claims that this is the view of causation held by Davidson, but it is much more widely held than that. There is much disagreement about what laws are, but little disagreement that laws underlie causal relations. Even among those who advocate counterfactual theories of causation, it is widely assumed that there

---

the fundamentality of counterfactual notions of causal relevance (Hitchcock 1995; Lewis 1973; Woodward 2003; Woodward and Hitchcock 2003). Ned Hall (2004) has recently argued for the ineliminability of two concepts of cause. In his essay he construes these as two sorts of causal relations between events, but much of his analysis of the difference between productivity and relevance carries over to this context.

[3] Or in virtue of certain facts about those events. Facts are more finely grained than events, but are generally related to properties. The fact that Bob coughed loudly is a fact about Bob's coughing event.

[4] (Fodor 1989, 64). Numbering of conditions has been changed.

are a set of laws of nature obtaining in this and other possible worlds, and that judgments about the nearness of a possible world to the actual world depend crucially upon these laws.

In my own work on mechanisms, I have suggested that what underlies most causal connections are not laws but mechanisms (Glennan 1996; 1997; 2002). I have argued that most laws – indeed all causal laws except the basic laws of physics – are really just descriptions of the behavior of mechanisms. I call such laws "mechanically explicable" for short. Thus, for instance, Mendel's laws are really just descriptions of various aspects of the mechanisms of sexual reproduction. Such laws are true only *ceteris paribus*, and the *ceteris paribus* clauses are, roughly speaking, violated in just those cases where the mechanism breaks down. For instance, Mendel's law of independent assortment is true only for genes located either on different chromosomes or far enough away on the same chromosome that recombination breaks linkage.

This view of higher-level laws has become more common, especially amongst philosophers of biology and psychology. Cummins puts it this way:

> Laws of psychology and geology are laws *in situ*, that is, laws that hold of a special kind of system because of its peculiar constitution and organization. The special sciences do not yield general laws of nature, but rather laws governing the special sorts of systems that are the proper objects of study. Laws *in situ* specify effects – regular behavioral patterns characteristic of a specific kind of mechanism (Cummins 2000, 121)

Some philosophers do not wish to call these generalizations laws. Woodward (2000) refers to them as invariant generalizations, while Craver (2007) has begun to call them mechanistically fragile generalizations. Schaffner (1993) sees these generalizations as parts of "theories of the middle range." But whether we choose to call them laws or something else, mechanists of all stripes agree that there are non-accidental, but not truly exceptionless generalizations that describe the behavior of mechanisms.

If this view of the relationship between laws and mechanisms is right, it suggests that Fodor's covering principle is true but misleading. It is true that wherever one event causes another, there will be a law appealing to properties of the events that says that the one event is *ceteris paribus* sufficient for the other[5]. The principle is misleading though because it suggests that the law is the truth maker for the causal claim. In fact, the mechanism is doing the causal work, and the law simply summarizes the behavior of the mechanism.

Cummins usefully makes essentially the same point when he says that laws in psychology and other special sciences specify effects. Since these laws in psychology and other special sciences describe the outward behavior of mechanisms, we don't explain anything by identifying a law; we have instead identified a target explanandum. The description of the mechanism responsible for the production of this effect provides the explanation (cf. Glennan 2002).

If laws aren't the truth makers for higher-level causal claims, then Fodor's covering principle doesn't really tell us what the role of properties is in making causal claims true. The mechanist's insight is that properties, and the nomological relations that hold between them, are not basic facts about the world, but are mechanically explicable

[5] To a second approximation, not all singular causal claims can be seen to be instances of laws. Laws are general, and there are sometimes singular causal claims involving essentially non-repeatable circumstances (say, e.g., claims about what caused Chamberlin to capitulate to Hitler at Munich) in which the events do not plausibly fall under general laws. See Glennan 2002 for further discussion.

facts. [6] In particular, the mechanist claims that when, in accordance with the canonical form, *c* causes *e* in virtue of properties *P*, *c* produces *e* via the operation of a mechanism, and the causally relevant properties are properties of the mechanism, its parts and organization, and its background operating conditions.

To get this account right, we must clarify the relationships between properties, mechanisms and laws. In particular, we must explore the consequences of the fact that a certain entity has certain properties is not a basic fact about that entity, but depends upon the nature and organization of that entity's parts. In the final two sections of this paper I will seek to spell this out by looking at two versions of the layered model. I will begin with Kim's functional model of reduction. I shall show how Kim's model involves a property bias which undermines his attempt to correctly characterize the relationship between properties at different layers. In the final section, I will offer an alternative account of higher-level properties that I hope both provides a clearer understanding of the relationship between properties and the layered model and more adequately addresses worries about epiphenomenalism.

## III. Kim and Higher-Level Properties and Causation

Jaegwon Kim has written perhaps more extensively than any other philosopher on the causal and explanatory role of higher-level properties. His principal concern has been the status of mental properties, but his model addresses the issue of higher-level properties generally. For the sake of definiteness, I will focus on the account he gives in

[6] Mechanists must come to terms with the fact that there appear to be some relations between fundamental constituents of the physical world whose causal relations are not mechanically explicable. My claims here and throughout this paper are meant to apply to the non-fundamental case (which includes just about everything we talk about in causal terms). For discussion about the relationship between fundamental and mechanical causes, see (Craver 2007, ch. 3; Glennan 1996; Glennan forthcoming; Psillos 2004)

his well-known *Mind in the Physical World*.[7] While Kim stakes out an extensive and nuanced set of positions on higher-level property causation, we can get a grip on the core of his position by considering the following figure:

$$\begin{tikzcd} M \arrow[r, dashed] & M^* \\ P \arrow[u] \arrow[r] & P^* \arrow[u] \end{tikzcd}$$

**Figure 1**

Suppose M and M* are mental or other higher-level properties, such that M causes M* to be instantiated. For instance M might be the property of seeing a tiger, and M* might be the property of being afraid. Minimal physicalism requires that mental properties supervene on physical properties, so let P and P* be the respective base properties for M and M*. Because M* supervenes on P*, the only way for M* to occur is for P to cause P*. P and P* are we'll presume neurological properties, and we have good reason to believe there is a complete story about how P causes P*. But, since P* necessitates M*, there is no genuine causal role for M in necessitating M*. Mental properties, because of their supervenience on physical properties, are epiphenomenal. This is what Kim calls the causal/explanatory exclusion argument, and it threatens mental and perhaps other higher-level properties with causal impotence (cf. Kim 38-47, 60-67). Kim argues that the only way out of this dilemma is reduction. By reducing mental properties to physical properties, they regain their causal powers. But in truth, it is not

7 (Kim 1998). Unqualified page number references are all to this work.

clear that one has really saved the causal powers of the mental. The reductionism Kim advocates skirts dangerously close to eliminativism.

One does not have to read a lot of his work to recognize that Kim conceptualizes reduction, supervenience and causation as relations between properties[8] Supervenience is characterized as a relationship between families of properties, and individual events stand in causal relations to each other only insofar as they are subsumed under laws, where laws are understood to be "objective connections between properties" (Kim [1984] 1993, 77). Mechanists should be suspicious of this flight from particulars. If laws are in fact only descriptions of the behavior of mechanisms, then it is the mechanisms rather than the laws that are doing the work. Similarly, when a higher-level entity has some property, there will be mechanisms that realize this property. The mechanistic point of view suggests that the properties of things and the causal relations that obtain between those things depend on mechanisms, and mechanisms are not just abstract relations between properties, but concrete collections of organized interacting entities. Failure to recognize this dependence on particulars is what I mean by property bias.

Kim thinks the only way out of the causal exclusion problem is to reduce higher-level properties to physical properties. To do this, he proposes what he calls the "functional model of reduction." The goal of this model is to show how an identity $M$=$P$ holds between higher-level properties and physical properties. If one can show that mental properties really are identical to physical properties then the causal exclusion problem doesn't arise. If, using the nomenclature of figure 1, $M$=$P$, then there is no puzzle about $M$ and $P$ both causing $M$* (=$P$*). Here is how Kim describes his model:

[8] See especially (Kim [1973] 1993)

> To reduce a property *M* to a domain of base properties, we must first "prime" *M* for reduction by construing, or reconstruing, it *relationally* or *extrinsically*. This turns *M* into a relational/extrinsic property. For functional reduction we construe *M* as a second-order property defined by its causal role—that is, by a causal specification *H* describing its (typical) causes and effects. So *M* is now the property of having a property with such-and-such causal potentials, and it turns out that property *P* is exactly the property that fits the causal specification (98).

As examples of functional reductions, Kim cites temperature, transparency, and the property of being a gene. In each case the higher-level properties are characterized in terms of their causal role. Temperature, for instance is a magnitude that will flow (in the sense that objects with different magnitudes that are put in contact will tend to equalize). Transparent objects will allow light to pass through. Genes will transmit characteristics between parents and offspring (cf. p. 25).

Kim's treatment of the relationship between higher-level and functional properties and their realizers reveals property bias. Kim's view seems to be that higher-level (and higher-order) properties are fixed *solely* by the base properties and their causal/nomological relations. He claims:

> Since whether or not *P* is a realizer of the functional property *M* is determined by the prevailing laws of nature, the realization relations remains invariant across all worlds with the same basic laws. Thus *M*=*P* holds in all nomologically possible worlds…. This result seems right: given the prevailing laws, DNA-molecules are the carriers of genetic

> information in this world, but in worlds with different basic laws, it may well be molecules of another kind that perform this causal work. (99-100).

There's something very wrong with saying that what can realize a functional property is entirely fixed by the "basic laws" of nature. It is true, to follow Kim's example, that were the basic laws of nature different, then DNA would not be able to function as a replicator, and hence as a realizer of the gene function. But it certainly does not follow from this that DNA is the only possible realizer of the gene function within the actual world. We know that RNA carries genetic information in some viruses. But beyond this, it is likely that in a world with the same physical laws, there could be a wide variety of molecules and other mechanisms that could code and replicate information in the way that our DNA-based genes do. Indeed, if we were to find life in distant parts of the universe we would be exceedingly surprised if the mechanisms of heredity in these life forms would closely resemble ours. Close similarity would be evidence for common descent.

The root of Kim's problem is his failure to recognize that the functional properties of complex entities depend not just on the basic laws of nature, but on the structure, organization and activities of the parts that make up these entities. You can not explain functions without talking about mechanisms. While Kim might not contest this point, offering a model of reduction in which realization is construed as a relation between disembodied properties shows a failure to recognize the metaphysical consequences of the particularity of mechanisms.

Let us return now to the functional model of reduction, and in particular to the question of whether it saves higher-level properties from causal impotence. Logically, functional/relational characterizations of these properties can all be treated as second order properties – having the property of being a property which does such and such. But if this is the right way to think about functional properties, Kim thinks (following (Block 1990)) that we should have serious doubts about the causal powers of functional and other second order properties. To cite one of Kim's examples, sleeping pills have the functional property of dormativity; that is, they have the property of being able to put people to sleep. Seconal and Valium are two pills with the property of dormativity, but what causes one to fall asleep is not the pills' dormativity, but the chemical properties of the pills that realize the dormativity. More generally, this suggests what Kim calls the "causal inheritance principle" (54). Any causal powers something has in virtue of its second- (or higher-) order properties are inherited from the first- (or lower-) order properties of that thing that realize those second (or higher-) order properties. Ultimately, the fact that functional properties introduce no new causal powers suggests that functional properties are not really novel properties at all, but are simply second-order ways of designating first-order properties. Dormativity is not a new property of sleeping pills over and above their chemical structures. It is simply a functional designator of a class of chemical structures which can play the causal role of inducing sleep.

This line of reasoning seems plausible, but when one considers that Kim offers functionalization as a general model of reduction, its implications for higher-level powers is pretty scary. Remember that Kim thinks that we can treat such properties as having a certain temperature and being a gene functionally, but if this is so, we appear to be robbing genes and even warm things of their causal powers! It is only the structural and intrinsic properties that realize functions that have the genuine causal powers, but every time we think we have an intrinsic property, we find that it in turn can be functionalized and reduced until we reach the fundamental physical level.

Do we even find instrinsic (as opposed to functional) properties at the physical level? Maybe not. If there were ever things that we would want to count as intrinsic properties, they would be basic properties of matter – like mass and charge. But what is it to have a certain mass or charge? The answer is given by basic laws of nature. What is it to have a certain mass if it is not, e.g., to accelerate towards other massive bodies in accordance with the law of universal gravitation and Newton's laws of motion? Kim endorses a principle (119), which he attributes to Samuel Alexander, that says in effect that the only reason for attributing reality to some thing is that it has causal powers. But to characterize a property in terms of its causal powers is to characterize it in terms of its second-order properties, and these properties don't have causal powers. Something is definitely wrong here.

Perhaps Kim has a solution to this dilemma. That solution involves us making a distinction between higher-level and higher-order properties. Higher-*order* properties are properties (or perhaps more properly property descriptions) obtained by quantifying over lower order properties. Higher-*level* properties are properties belonging to aggregates that do not belong to their components. Orders are a logical notion, while levels are mereological. Kim suggests that there are no problems with attributing new causal powers to higher-level properties. For instance, if a radio weighs 1 kg, that weight gives it certain causal powers-- e.g., the ability to stretch a spring by some distance *k*-- that would not belong to any of the radio's parts. Moreover, it will generally be the case that new causal powers are introduced not just by piling together parts but by organizing them into mechanisms. A radio has the power to receive a signal and convert it into a sound, but it only has this power in virtue of the functional organization of the parts. A pile of radio parts is not a radio (*cf*. Craver 2007, ch. 5).

Kim thinks of these new causal powers as being attributed to what he calls micro-based properties:

> $P$ is a micro-based property just in case $P$ is the property of being completely decomposable into non-overlapping proper parts $a_1$, $a_2$, … $a_n$ such that $P_1(a_1)$, $P_2(a_2)$, … $P_n(a_n)$, and $R(a_1, a_2, \ldots a_n)$ (84).

Kim cites as an example of a micro-based property the property of being water. Water molecules have three atomic parts that stand in a particular bonding relation to each other. Kim emphasizes that micro-based properties are not functional properties but are the realizers of functional properties. We can see for instance that the micro-based property of the radio is the structural property of the radio that realizes the functional property of being something that converts radio signals to audio signals. Kim's account of these properties makes explicit that an entity $x$ has such a property $P$ in virtue of the kinds of parts of which it is composed and the relations that obtain between its parts. This goes some way towards correcting property bias and moving towards mechanisms

Although the notation is far from perspicuous, it might seem possible for Kim to pack enough information into the micro properties $P_i$ and their relation $R$ to provide an account of the organization of the mechanism responsible for the realization of $R$. It does go some way, because it shows that the properties of things depend upon the micro-structure of those things, but there is more to mechanism than structure.

The microstructure of a mechanism is what allows its parts to interact in order to produce the behavior of the thing, or as we would say in this instance, to realize its properties, but an account of how the mechanism works requires us to describe how things with that structure do in fact act and interact to produce the effect. One of the functional properties of water, for instance, is that it is a solvent. That water is a solvent depends upon its structure, but to describe the structure of water is not in itself to explain its solvent properties. A proper description of the solvent properties of water involves explaining the mechanism by which water molecules interact with molecules of a dissolving substance like salt. Note also that we can formulate laws to the effect that if *x* is a volume of water and *y* is a volume of salt, then, given appropriate volumes, the water will dissolve the salt. But this law is, as Cummins says, an *effect* – a description of what happens that stands in need of explanation. The effect is explained by describing the mechanism by which water molecules break the ionic bonds in salt.

Kim suggests to us that we should conceive of micro-based properties as higher-level properties, because these are properties that can only be had by things which are themselves composed of proper parts. These micro-based properties will in general have causal powers that none of their parts have on their own. While I certainly would not deny that organized aggregates have novel causal powers, a micro-based property is not thereby a genuine higher-level property. If we describe some thing *x* in terms of its micro-structure, the micro-based property *P* is taken to be identical to a complex conjunction of properties of and relations between its parts. If having the property really amounts just to having that micro-structure, then this is a complex low-level property. Suppose for instance we have a micro-based property of the brain which serves as a realizer of the functional property of being in pain. Ultimately, via substitution, this property is going to be shown to be an exceedingly complicated structural description of the various neurological components of the brain which are involved in the realization of this function. Kim would have us believe that this complex conjunction of properties of and relations between microscopic constituents of the brain is all we could mean by a higher-level property. The problem with this account is that it makes the identity conditions for having a property depend upon the exact microstructure of the complex system. But systems with distinct micro-based properties may instantiate the same higher-level property.

To see this point more clearly, let's consider a very simple example of the relationship between micro-properties and macro-properties. Consider a specific shape property, the property of being a three inch cube. Is this a micro-based property? Well, if we consider any particular three inch cube, it would appear to be so. The cube would be made of parts, perhaps small solid shapes, or perhaps the cube would be made of a uniform material, so that its constituents might be molecules in some lattice structure. For simplicity's sake, imagine that we are building three inch cubes out of a block set made of 3"x 1"x 1" rectangular blocks. There are, ignoring possibilities introduced by rotations, three ways to put these blocks into a three inch cube, as illustrated in figure 2.

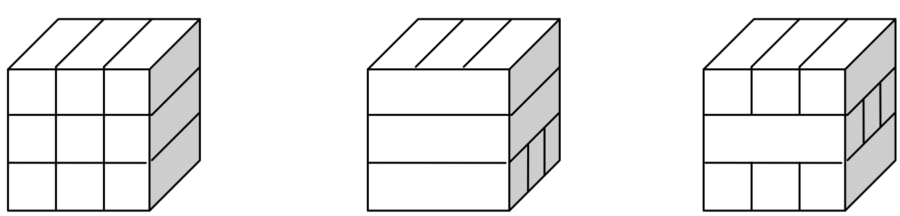

Figure 2: Realizations of a three inch cube

For any particular cube, we will be able to specify which blocks were used and how they were stacked together, and this will be a micro-based property in Kim's sense, but the property of being a three inch cube cannot be identified with any particular one of these properties. If we are to identify being a cube with a micro-based property, that property is going to be disjunctive. But, says Kim and most other philosophers, disjunctive properties aren't really properties at all (106-108). Certainly they don't have causal powers.

Kim's attempt to characterize higher-level properties as micro-based properties thus appears to discount properties like shape as genuine structural properties, or at least to rob them of causal powers. The most obvious way for Kim to respond to this problem is to stop treating shape as a structural property and functionalize it. Shape ceases to be a higher-level structural property and becomes a higher-order functional property. It is of course possible to characterize a property like shape in terms of causal dispositions. For instance, we might say that two keys have the same shape if they can fit in the same locks. But this is a desperate measure that I would not recommend. If you must functionalize shape, it seems likely that you must functionalize all higher-level properties. This move will rob higher-level properties of their causal powers, since causal powers belong according to Kim to the realizer.

The real difficulty with functionalizing all of these properties is that it will cause one to lose the scientifically invaluable distinction between structure and function. To see the consequences of this move, let's consider one further case in which two different shapes (structures) can realize the same function. Dual-keyed locks are locks that can be opened either by a standard key or a master key. Thus, keys of either shape are realizers of the function of being a lock opener. We certainly would not want to attribute to these keys the causal power to open locks in virtue of their properties of being lock openers, but we would want to attribute causal powers to the keys in virtue of their shape. And when we have two copies of the standard key, it is in virtue of the same higher-level structural property (namely their identical shape) that they realize the functional property of being a door opener, but when we have a master key and a standard key, they realize the functional property in virtue of a different higher-level property. An account of the relation between structural and functional properties which explains all realization relations in terms of micro-structure is going to miss the real causal powers of macro-structural properties.

## IV. Mechanisms and higher level causes

In closing I would like to propose an alternative solution to the causal exclusion problem that appeals to the mechanistic approach to causation I've advocated, as well as to the distinction between productivity and relevance introduced in the first section of this paper. Let us apply the canonical form of a causal claim to the case of a key opening a lock. Instantiating (C) in this case we get:

> The turning of the key causes the opening of the lock in virtue of the shape and rigidity of the key that turned.

The productive causal relation holds between one event—the key turning – and another event – the lock opening. Equivalently we could say that the key – the object implicated in the key-turning event – produced the lock opening. To say this is to say simply that the key is a working part of the mechanism, and that key turning is one of its activities. What we should not say is that certain *properties* are productive.[9]

Suppose key *k* opens the lock. Let us say that it instantiates a certain microstructure *M* and a certain macrostructure (shape and rigidity) *S*. The key *k* is thus an instance of both *S* and *M*. If we think of productivity as a property of properties then we will mistakenly ask the question of whether it is *S* or *M* that produces the opening. But the very same key is an instance of both properties. It is the key with microstructure *M* and macrostructure *S* that opens the lock. To ask whether it was *S* or *M* that *produced* the opening involves a category mistake of attributing productivity to properties.

When we wonder whether it is the microstructure or the macrostructure that matters, we are really asking a question about relevance. Intuitively it should be clear that it is the macrostructure – the shape and rigidity – of the key that are the relevant properties, but let us consider what informs that intuition.

Carl Craver (2007, ch. 6), in his attempt to defend the relevance of higher-level properties has suggested that we should understand relevance by appealing to the sorts of experimental manipulations we use to determine relevance. Roughly, he suggests that we determine the relevance of a property by performing experiments to see what happens when we change that property. This procedure will in fact ground our intuitions in this case. We can demonstrate that shape and rigidity are the relevant higher-level properties by varying the shape and rigidity of keys which we put in the lock. Keys with different shapes won't unlock the lock, nor will keys made of non-rigid materials. Equally we can

[9] Jackson and Petit (1990) offer a somewhat similar solution to the one I'm proposing here, when they argue that higher level properties are causally relevant, while lower-level properties are the only ones that are causally efficacious. Efficacy plays much the same role in their account as does productivity in mine. Their mistake in my view was to think of efficacy as a property of properties. As I've argued earlier in this paper, it is only objects realizing properties rather than the properties themselves that can produce anything.

demonstrate that microstructural properties are irrelevant by showing how rigid keys of the same shape will all work to open the lock, despite substantial variations in, e.g., the materials of which the keys are made.

Experiments of this sort are clearly crucial to our determinations of causal relevance, but I would not want us to suppose that these epistemic procedures are *constitutive* of causal relevance. Experiments help us to understand the organization of the mechanisms that determine the causal relationships between events, but it is the organization and operation of the mechanisms themselves, not the experiments, that make certain properties causally relevant. Consider again our lock and key. Cylinder locks have a mechanism involving a number of parts – most notably a cylinder in a casing with a slot for the key, and a set of pins that fall down into the cylinder and can be pushed up into the casing by the key. Key shape is causally relevant because it determines alignment of the pins, and proper alignment of the pins is what allows the cylinder to turn. Anyone can establish that key shape is a relevant property, simply by doing experiments with different shaped keys. But we do not understand why shape is relevant until we understand how the lock works. By understanding the parts and organization of the mechanism we are able to explain and predict why certain keys will work and others will not.

More generally, if we have a well-understood mechanism that is productive of some behavior or effect, that model will identify certain properties of the parts (or perhaps of collections of the parts) that determine how those parts interact. These properties will figure in any generalizations we have concerning interactions of parts of the mechanism, so we will know that if these properties were to change, then the behavior of the parts, and hence of the mechanism would change. A description of the organization and operations of the mechanism will thus identify the relevant properties.

Simple examples like the lock and key allow us to understand how the mechanistic model can be used to defend the reality and causal relevance of higher-level properties against the challenges posed by the exclusion argument. Much of the real promise of the mechanistic approach, however, lies in explaining more complex causal processes. The sorts of processes studied in the life and social sciences typically are hierarchical and involve mechanisms in which there is a complex causal interdependence between part and system level properties. For instance, in the mechanism of human thermoregulation, changes in ambient temperature lead to changes in behaviors of thermoreceptor cells which in turn lead to changes in the hypothalamus, which in turn lead to both involuntary low-level responses like constriction of blood vessels and high-level voluntary processes like putting on a sweater.

In a recent paper, Craver and Bechtel have provided an analysis of how so-called top-down causation can be explained and demystified within a framework of hierarchical mechanistic analysis. Their view is that supposed cases of top-down and bottom-up causation are not really cases of causation as such, but are what they call "mechanistically mediated effects." As they put it, "[these] effects are hybrids of constitutive and causal relations in a mechanism, where the constitutive relations are interlevel, and the causal relations are exclusively intralevel (Craver and Bechtel 2007, 547). The idea here is that higher-level systemic properties are not caused by the interactions of its parts, but are constituted by them. Consider, for instance, a case in which ambient temperature drops, causing triggering of the body's thermostat in the hypothalamus. This might look like a case of a property of the whole body (its ambient temperature) causes a change in one of its parts (signaling from the hypothalamus). On closer analysis, however, what is going on at the neurological level is lots of local changes to the environment of individual thermoreceptor cells which, via some intracellular mechanism, trigger firings that collectively reach a threshold required to trigger negative feedback mechanisms controlled in the hypothalamus. Change in the ambient temperature is not a high-level event over and above changes in the thermal energy in regions surrounding the thermoreceptor cells. Changes in the ambient temperature don’t cause energetic changes in these regions; they are constituted by them.

Craver and Bechtel’s analysis fits together nicely with the account I have offered of the distinction between productivity and relevance. A change in ambient temperature is an event which has both microstructural and macrostructural properties. For reasons I suggested in the lock and key case, it is a mistake to ask which of these properties produces changes in the hypothalamus. My reasons are essentially the same as those offered by Craver and Bechtel. Changes in microstructure and changes in macrostructure are not distinct events that might stand in productive causal relations. But at the same time, we can save the intuition that causation is top-down, by noting that in a case like thermoregulation, it is changes in the high-level properties that are causally relevant. As we’ve noted in this case, triggering of the negative feedback mechanism in the hypothalamus is a threshold effect that depends upon the overall volume of signals from thermoreceptors throughout the body. Variations in which receptors are firing that do not change the overall signal volume will not change the system’s overall behavior. For this reason, the sorts of manipulations that will bring about triggering of the negative feedback mechanism will always involve changing a high-level property of the system.

What I've tried to suggest in this paper is that the philosopher's traditional ontological categories of objects, properties and laws is simply not nuanced enough to explain the nature of causal relations in the layered world. It is not that there aren't objects, properties and laws in this world. Of course there are. It is just that these ontological categories are not basic, and when we treat them as such we get a distorted picture of the nature of causal relations. One big payoff of the mechanistic approach is that it can tell us more about what objects, properties and laws really are. In so doing, we get a much improved account of how the higher-level entities and properties that are so central to both our scientific and everyday discourse can matter in the world.

## Works Cited


Bennett, Jonathan Francis (1988), *Events and their Names.* Indianapolis: Hackett Publishing.

Block, Ned. (1990), "Can the Mind Change the World?" in George Boolos (ed.), *Meaning and Method: Essays Honor of Hilary Putnam.* New York: Cambridge University Press.

Bogen, Jim (2004), "Analysing Causality: The Opposite of Counterfactual is Factual," *International Studies in the Philosophy of Science* 18(1): 3-26.

Craver, Carl F. (2007), *Explaining the Brain: What a Science of the Mind-Brain could be.* New York: Oxford.

Craver, Carl and William Bechtel (2007), "Top-down Causation without Top-down Causes," *Biology and Philosophy* 22: 547-563.

Cummins, Robert. (2000), "'How does it Work?' vs. 'What are the Laws?' Two Conceptions of Psychological Explanation." in F. Keil and R. Wilson (eds.), *Explanation and Cognition.* Cambridge, MA: MIT Press.

Dowe, Phil (2000), *Physical Causation.* Cambridge Studies in Probability, Induction, and Decision Theory. New York: Cambridge University Press.

Fodor, Jerry (1989), "Making Mind Matter More," *Philosophical Topics* 17(1): 59-74.

Glennan, Stuart S. (1996), "Mechanisms and the Nature of Causation," *Erkenntnis* 44(1): 49-71.

———. (1997), "Capacities, Universality, and Singularity," *Philosophy of Science* 64(4): 605-626.

———. (2002), "Rethinking Mechanistic Explanation," *Philosophy of Science* 69(3 Supplement): S342-S353.

———. (forthcoming), "Mechanisms," in Helen Beebee, Christopher Hitchcock and Peter Menzies(eds.), *Oxford Handbook of Causation.* New York: Oxford University Press.

Hall, Ned. (2004), "Two Concepts of Causation," in John Collins, Ned Hall and L. A. Paul (eds.), *Causation and Counterfactuals.* Cambridge, MA: MIT Press.

Hitchcock, Christopher Read (1995), "Discussion: Salmon on Explanatory Relevance," *Philosophy of Science* 62(2): 304-320.

Jackson, Frank, and Philip Pettit (1990), "Program Explanation: A General Perspective," *Analysis* 50(2): 107-117.

Kim, Jaegwon ([1973] 1993) "Causation, Nomic Subsumption, and the Concept of Event," in Kim 1993, 3-32. Originally published in *Journal of Philosophy* 70: 217-236.

———. ([1984] 1993) "Concepts of Supervenience," in Kim 1993, 53-78. Originally published in *Philosophy and Phenomenological Research* 45 (December): 153-176.

———. (1993), *Supervenience and Mind: Selected Philosophical Essays.* Cambridge studies in philosophy. New York: Cambridge University Press.

———. (1998), *Mind in a Physical World : An Essay on the Mind-Body Problem and Mental Causation.* Cambridge, MA: MIT Press.

Lewis, David (1973), "Causation," *Journal of Philosophy* 70: 556-567.

Machamer, Peter, Lindley Darden, and Carl F. Craver (2000), "Thinking about Mechanisms," *Philosophy of Science* 67(1): 1-25.

Psillos, Stathis (2004), "A Glimpse of the Secret Connexion: Harmonizing Mechanisms with Counterfactuals," *Perspectives on Science* 12(3): 288-319.

Salmon, Wesley C. (1984), *Scientific Explanation and the Causal Structure of the World.* Princeton, N.J.: Princeton University Press.

Schaffer, Jonathan. (2003), "The Metaphysics of Causation," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy.* (Spring 2003 Edition) URL = <http://plato.stanford.edu/archives/spr2003/entries/causation-metaphysics/>.

Schaffner, Kenneth (1993), *Discovery and Explanation in Science and Medicine.* Chicago: University of Chicago Press.

Woodward, James (2000), "Explanation and Invariance in the Special Sciences," *British Journal for the Philosophy of Science* 51(2): 197-254.

———. (2003), *Making Things Happen: A Theory of Causal Explanation.* Oxford: Oxford University Press.

Woodward, James, and Christopher Hitchcock (2003), "Explanatory Generalizations, Part I: A Counterfactual Account," *Noûs* 37(1): 1-24.