Biology and Philosophy (2007) 22:547–563
DOI 10.1007/s10539-006-9028-8


# Top-down causation without top-down causes

CARL F. CRAVER[1,*] and WILLIAM BECHTEL[2]
[1]*Philosophy-Neuroscience-Psychology Program, Washington University, One Brookings Hall, St. Louis, 63130 MO, USA* [2]*Department of Philosophy-0119, University of California, 9500 Gilman Drive, LaJolla, San Diego, CA, 92093 USA; *Author for correspondence (e-mail: ccraver @artsci.wustl.edu)*

Received 22 December 2005; accepted in revised form 10 March 2006

**Key words:** Levels, Mechanisms, Top-down causation, Interlevel causation, Constitution, Explanation, Emergence, Reduction

**Abstract.** We argue that intelligible appeals to interlevel causes (top-down and bottom-up) can be understood, without remainder, as appeals to mechanistically mediated effects. Mechanistically mediated effects are hybrids of causal and constitutive relations, where the causal relations are exclusively intralevel. The idea of causation would have to stretch to the breaking point to accommodate interlevel causes. The notion of a mechanistically mediated effect is preferable because it can do all of the required work without appealing to mysterious interlevel causes. When interlevel causes can be translated into mechanistically mediated effects, the posited relationship is intelligible and should raise no special philosophical objections. When they cannot, they are suspect.

## Introduction

Many philosophers (e.g., Alexander 1927; and several authors in Andersen et al. 2000) and scientists (e.g., Morgan 1927; Campbell 1974; Sperry 1976) appeal to top-down causes in their explanations. Such appeals evoke concerns that the notion of top-down causation is incoherent or that it involves spooky forces exerted by wholes upon their components. In our view, the phrase 'top-down causation' is often used to describe a perfectly coherent and familiar relationship between the activities of wholes and the behaviors of their components, but the relationship is not a *causal* relationship. Likewise, the phrase 'bottom-up causation' does not, properly speaking, pick out a causal relationship. Rather, in unobjectionable cases both phrases describe *mechanistically mediated effects*. Mechanistically mediated effects are hybrids of constitutive and causal relations in a mechanism, where the constitutive relations are interlevel, and the causal relations are exclusively intralevel. Appeal to top-down causation seems spooky or incoherent when it cannot be explicated in terms of mechanistically mediated effects.

To focus on the notion of top-down causation, we assume both that there are higher level causes and, further, that all higher-level causes are fully explained by constitutive mechanisms.[1] In assuming that there are higher-level causes, we mean to assume only that

(1) mechanisms are organized collections of entities and activities,
(2) mechanisms are affected by (and have effects upon) things
(3) the parts of the mechanism taken singularly cannot be so affected or have such effects.[2]

Ample evidence for claims 2 and 3 can be found in the ability of engineers to construct new devices from old parts and in the ability of natural selection to develop novel solutions to environmental challenges by rearranging old parts or inserting new parts into an old mechanism. However, our intent is not to argue for the existence of higher-level causes but to assume they exist in order to explore the possibility of top-down causation. Top-down causation is obviously not possible if higher-level causes do not exist. We also assume (but will not argue) that all higher-level causes are explained by lower-level mechanisms. We demonstrate that there can be legitimate appeals to top-down 'causation' (more properly, mechanistically mediated effects) even if these assumptions hold.

## Levels

To say that a causal relation is bottom-up or top-down is to say that things at one level are causally related to things at another level.[3] The term 'level' plays many roles in science. There are levels of abstraction, being, causation, description, explanation, function, and generality, to name a few, and these are not the same. For each, there is a different sense in which a cause can be said to be at the top (or bottom) and a different sense in which its influence is propagated downward (or upward). In this discussion, we focus on levels of mechanisms.

In levels of mechanisms (see Craver 2001; forthcoming), an item X is at a lower level than an item S if and only if X is a component in the mechanism for some activity $\psi$ of S. X is a component in a mechanism if and only if it is one of the entities or activities organized such that S $\psi$s. For that is what mechanisms

[1] Compare assumptions (1) and (2) in Kim's (2000, 302) discussion of top-down causation.

[2] We do not assume that the mechanism has causal powers over and above the organized collection of their parts. This claim faces much-discussed challenges that do not concern us here.

[3] We are intentionally noncommittal about the relata in the causal relationship, and will understand them as properties, events, processes, and activities as appropriate in context. The problems that we shall discuss can be formulated in each of these categories.

are: they are entities and activities organized such that they exhibit a phenomenon.[4] Scientists discover lower levels by decomposing the behavior of a mechanism into the behaviors of its component parts, decomposing the behaviors of the parts into the behaviors of their parts, and so on.

The eye is a familiar and unproblematic example of a multilevel mechanism. At the highest level, the eye transduces light into a pattern of neural activities in the optic nerve. This process can be decomposed into lower-level components and their activities. The light enters the eye, it is inverted and focused by a lens, and it is projected onto the retina, where the information in the light is converted into a pattern of neural activity in the optic nerve. The conversion of light into patterns of neural activity by the retina can itself be decomposed into different components: in particular, the rods and cones that change their electrical state depending on specific features of the light stimulus (such as wavelength and intensity). Another level down, rod cell activation is also sustained by a mechanism. Light is absorbed by and activates rhodopsin, which then stimulates G-proteins. These G-proteins activate cyclic GMP phosphodiesterase, which catalyzes the conversion of cyclic GMP to 5′-GMP. Lowering the concentration of cyclic GMP causes sodium channels to close, reducing the inward sodium current and thereby hyperpolarizing the cell (see Kandel et al. 1991). Each new decomposition of a mechanism into its component parts reveals another lower-level mechanism until the mechanism bottoms out in items for which mechanistic decomposition is no longer possible.[5]

Though not the only sense of level that might be relevant to discussions of interlevel causation, levels of mechanisms are unquestionably a (if not the) central sense of levels in such discussions (see, e.g., Wimsatt 1976; Churchland and Sejnowski 1992; Kim 2000). Levels of mechanisms are literally ubiquitous in the biological sciences and, as our examples illustrate, these are the kinds of cases in which debates about the possibility of interlevel causation arise. Furthermore, debates about reduction and reductionism in biology are sometimes (and, if we are right, misleadingly) cast as debates about whether the direction of causation runs exclusively from lower-level mechanisms to higher-level phenomena or whether higher-level phenomena act back upon lower-level mechanisms.[6]

A final reason for our focus on levels of mechanisms is that they have many of the features typically associated with levels in discussions of top-down causation. Because mechanisms are organized collections of components and

[4] For related accounts of mechanisms, see: Bechtel and Richardson 1993; Bechtel 2006; Craver forthcoming; Glennan 1996; Machamer et al. 2000; Woodward 2003. The differences among these views are of no consequence for the present discussion.

[5] In the practice of scientific explanation, decomposition in a given inquiry bottoms out when the investigation reaches entities and activities that are viewed as unproblematic or for which investigators lack tools for further decomposition. Few decompositions extend more than a small number of levels.

[6] Our view can accommodate strong, medium and weak downward causation in the sense of Emmeche et al. 2000.

their activities, no component can be larger than the mechanism as a whole, and so levels of mechanisms are ordered by size. For analogous reasons, higher-level behaviors act over longer time-periods than lower-level activities. The absorption of light by rhodopsin necessarily happens faster than the same light can cause the rod to hyperpolarize. Most fundamentally, levels of mechanisms are a species of compositional, or part-whole, relations. In contemporary debates about reduction and interlevel causation, it is common for authors to talk about 'levels of aggregation,' 'levels of organization,' 'levels of complexity,' and 'mereological levels.' Such descriptions apply to levels of mechanisms as well. Higher levels of mechanisms are *aggregated* (i.e., built up from) or *composed* from *parts* that are *organized* into more *complex* spatial, temporal, and causal relations.

Our decision to focus on levels of mechanisms allows us to clear up three possible sources of confusion in thinking about interlevel causation.

First, the focus on levels of mechanisms separates the question of interlevel causation from the question of whether there can be causal relationships among items of different sizes. We are unaware of any consideration that would prevent one from allowing that large things (even very large things) sometimes interact with small things (even very small things). Viruses infect elephants. Cocaine leaves its mark on cells, human beings, and societies. Viruses are not parts of cells. Cocaine is not part of the human organism. None of these examples, that is, involves levels of mechanisms. Although levels of mechanisms are ordered from the smallest items at the bottom to largest items at the top, this ordering is derivative upon the constitutive relationship. As we argue in the next section, it is the constitutive relationship among levels that makes talk of top-down causation seem problematic.

Second, levels of mechanisms are not monolithic divides across all of nature. Levels of mechanisms are defined locally, within the context of a given type of mechanism. One is thinking of levels as monolithic divides across all of nature when one thinks of levels as levels of sciences (e.g., economics, psychology, biology, chemistry, physics; see Oppenheim and Putnam 1958) or as levels of entities (e.g., societies, individuals, organs, molecules, atoms). Again, we find no metaphysical puzzle imagining that items in the proper domain of one science, however that domain is defined, interact with items in the proper domain of another science. Nor is there a clear reason why items in one ontological category (e.g., the molecules) cannot interact with those in another ontological category (e.g., the cells). Again, the primary difficulty envisioning interlevel causation seems to arise from the constitutive relationship between a whole and its parts and not from the appearance of wholes and parts in different theories or from the inclusion of wholes and parts in different ontological categories.

Finally, higher levels of mechanisms are, by definition, mechanistically explicable. One might object that we thereby exclude 'emergent' causes by fiat.

A defining mark of 'strongly emergent properties' is that *they have no mechanistic explanation*.[7] The organization of components in a mechanism may allow the novel property to 'emerge,' but the property has no explanation in terms of the operation of that mechanism. We acknowledge that there can be no levels of mechanisms when decomposition is impossible in principle. We draw two conclusions from this observation. First, the notion of 'level' involved in considering cases of emergence is not the same as the notion of level that is so ubiquitous in biology. Levels of mechanisms are constitutive levels; levels of strong emergence are not. For this reason, the notion of strong emergence can borrow no legitimacy from its loose association with the levels of mechanisms so ubiquitous in biology and elsewhere. Second, our account places a burden on the defender of strongly emergent properties to explain why top-down causation from emergent to nonemergent properties is different from mundane causation between two distinct properties. For this reason, it seems to us best to separate the question of whether strongly emergent properties are possible (in some sense of the word possible) from the question of whether top-down causation is possible. The problem of top-down causation can arise without strong emergence, and the possibility of strongly emergent properties does nothing to make that problem more difficult to solve (indeed, if our second conclusion is correct, it can make it considerably easier to solve).

There may be other senses of level, and there may be other notions of interlevel causation, but this paper is about interlevel causes in levels of mechanisms. Once we are clear on how to think about top-down causation in levels of mechanisms, we can then ask if there are other compelling problems that deserve this name.

## Interlevel causes

Many common assumptions about the nature of causation preclude the possibility of causal relations between parts (components) and wholes (mechanisms). To start with an especially clear example, consider the view that all causation involves transmitting something such as a mark (Salmon 1984) or a conserved quantity (Dowe 2000) from one event, object, or process to another.[8] This is a minority view to be sure, but the core idea in these accounts is implicit in many of the metaphors used to describe causal connections. Causation is frequently described as a kind of cement, glue, spring, string, or some other physical transmission or exchange from one object, process or event to another through contact action or through a propagated signal (see Hitchcock 2003).

[7] Others (see Boogerd et al. 2005) use 'strong emergence' for properties of complex wholes which cannot be predicted from knowledge of the parts in isolation or in simpler collections or parts. Such 'strongly emergent' properties are mechanistically explicable and hence not incompatible with our account of levels.

[8] Dowe (2000) intends only to offer an account of physical causation, not causation at higher levels of organization.

Such a conception of causation as a physical connection between two things does not accommodate interlevel causes between mechanisms and their components because mechanisms and their components are not distinct events, objects, or processes. Given the compositional relations between mechanisms and their components, the space-time path of the mechanism includes the space-time path of its components. They coexist with one another, and so there is no possibility of their *coming to* spatiotemporally intersect with one another. If a conserved quantity is possessed by one of the components (say, a certain mass or a charge), that conserved quantity is also possessed by the whole. If one of the parts bears a mark, that mark is always already born by the whole (by virtue of being born by its parts). The marks do not need to be transmitted upward or downward to have their 'effects;' their effects are inherited constitutively, not causally.

The constitutive relation between mechanisms and their components conflicts with many other common assumptions about causation. Most theories of causation, at least since Hume, have assumed that causes and effects must be wholly distinct. Lewis is explicit:

> C and E must be distinct events – and distinct not only in the sense of nonidentity but also in the sense of nonoverlap and nonimplication. It won't do to say that my speaking this sentence causes my speaking this sentence or that my speaking the whole of it causes my speaking the first half of it; or that my speaking causes my speaking it loudly, or vice versa. (2000, 78)

The problem can be illustrated just as easily with examples drawn from science, such as the example of the eye introduced above. The change in the conformation of rhodopsin is a stage in the signal transduction pathway. Although the change in rhodopsin is a cause of the electrical signal generated by the cell, scientists do not regard it as a *cause* of signal transduction. Of course, there is no stopping someone who wants to use the word 'cause' in this way, but to do so is to expand the extension of 'cause' to cover relationships that are already characterized without remainder by the word 'component.'

Many theories of causation assume that causes precede their effects. This feature of causation is often disputed (see Faye 2005), and some accounts of causation (e.g., Reichenbach 1958) are designed as the foundation of an account of the temporal order, and so do not assume the temporal asymmetry of causation. Nonetheless, there is a problem lurking in the temporal relationship between the relata in interlevel causes. As Kim (2000) argues, the possibility of bottom-up and top-down influence 'propagated' simultaneously across levels results in problematic causal circles. For example, one might believe that if an object, X, has its causal powers in virtue of possessing a property, P, then if X is to exercise its powers at time $t$, X must possess P at $t$. And one might believe further that if something causes X to acquire P at $t$, then

x does not already possess P at $t$ until that something has acted. If X's acquiring P at $t$ is a cause of S's having $\psi$ at $t$, and S's having $\psi$ at $t$ is a cause of X's having P at $t$, then it appears that X's acquiring P at $t$ cannot cause S to have $\psi$ until S's having $\psi$ causes X to acquire P. In that case, it is little wonder that talk of interlevel causation strikes us as mysterious.

To avoid this problem, one might assume that causal transactions across levels take time: the effects of changes to a component alter the behavior of the mechanism as a whole at some later time, and vice versa. This would ameliorate some of the worry about the temporal order of causation, but it raises a related worry about the asymmetry of causation.[9] It is a widely accepted condition on accounts of causation that they account for the asymmetry of causal dependency. The sun's elevation causes the length of the shadow, but the length of the shadow does not determine the elevation of the sun. The virus produces the spots on the skin, but the spots on the skin do not cause the infection with the virus. Causes produce their effects, and (at least in many cases) *not* vice versa. Examples such as these have the staying power that they do because the asymmetry of causation is so fundamental to our very idea of causation. While at least some cases of intralevel causation are asymmetrical, all of the interesting cases of interlevel causation are symmetrical: components act as they do because of factors acting on mechanisms, and mechanisms act as they do because of the activities of their lower-level components.

One consideration that makes invocation of interlevel causal discourse so tempting in the context of mechanisms is the way scientists experiment on mechanisms. Experiments designed to test whether a given component is part of a mechanism have much the same structure as those used to test causal relations at the same level (Craver 2002, forthcoming; Bechtel in press). One strategy scientists employ is to seek out correlations between the presence of some component X or the occurrence of one of its activities $\phi$ and the behavior $\psi$ of the mechanism as a whole S. The correlation between the number of microsomes in a cell's cytoplasm and its activity of secreting chemical products was a clue that microsomes should be included as part of the mechanism of protein synthesis (whereas the number of other cell structures, such as mitochondria, was not so correlated, and thus they were not included as part of that mechanism). Lesion methods are often used to argue that a component is part of a mechanism: deficits in the mechanism S's behavior as a whole $\psi$ after removing or retarding some of the activities $\phi$ of some component X is prima facie evidence that X is a component in the mechanism for S's $\psi$-ing.[10] Stimulation experiments (in which one stimulates some X to $\phi$ and then observes the S's $\psi$-ing) are also used to argue that a given item is a component in the mechanism. The ability to stimulate or regulate biochemical pathways through

[9] Again, this principle has been questioned. See Price (1996) for a lengthy review.

[10] It is only prima facie since if the component is not interactively integrated into the system, it could also be construed as constituting a background condition on the operation of the mechanism, not as part of it.

the addition of precursors or catalysts is frequently used as an argument for the relevance of the added stimulant; and neuroscientists frequently use pharmacological agonists in their search for mechanisms in the brain. Finally, one can intervene to change the behavior $\psi$ of the mechanism as a whole, and observe the consequences for the behaviors $\phi$'s of the components. An example is neuroimaging (i.e., functional uses of PET and MRI scanning), in which correlations between blood flow and cognitive tasks are used to determine which brain regions contribute to the cognitive system.[11] These four kinds of experiment share a form with Mill's methods, giving the sense that a *causal* relationship is being tested.

Such experimental techniques involve either intervening to alter a component of the mechanism and observing the behavior of the mechanism as a whole or intervening to alter the behavior of the mechanism as a whole and observing the behaviors of one or more of its parts. In using both sorts of strategies, or in simply seeking correlations between the presence of components and the behavior of whole mechanisms, scientists seem to be exploiting a symmetrical relation. Yet, typically experimental techniques for testing causal claims work from cause to effect and not (generally) vice versa. One cannot generally intervene to change the effect and thereby alter the cause. Those who endorse the asymmetry of causation will thus have another reason to find interlevel causation mysterious.

These objections to interlevel causation might seem pointless if there were no positive alternative account of the phenomena for which interlevel causation is often invoked. Our account of mechanistically mediated effects, however, provides just the alternative account that is needed: it allows us to retain a univocal conception of causation as intralevel and to account for relations between levels in terms of constitution. On this account, the symmetry of the interlevel relationship and the techniques employed to investigate it is readily explained. The relation is symmetrical precisely because the mechanism as a whole is fully constituted by the organized activities of its parts: a change in the parts is manifest as a change in the mechanism as a whole, and a change in the mechanism is also a change in at least some of its component parts. There is no need to extend the word 'causation' to cover cases of this sort so long as one can describe the putative case of interlevel causation without remainder without doing so.

None of these arguments is decisive, or course. One wishing to make room for interlevel causation may wish to fashion the notion of causation in such a way as to avoid all of the potential obstacles we have mentioned. Alternatively, one might respond that there is no univocal notion of causation from which the assessment of interlevel causes can take place. However, we hope to have shown that to accept interlevel relationships as casual violates many of the central ideas associated with the concept of causation. Patricia Churchland

[11] Again, these judgments are prima facie ones and depend on discovering that the components are deeply integrated into the system.

(1993) expresses this negative reaction to interlevel causation with the help of an example from the *Betty Crocker Cookbook*. Betty correctly explains that microwaves work by accelerating the component molecules in the food. However, she takes a decidedly wrong turn when she explains further that the excited molecules rub against one another and *generate* heat through friction. Betty's error, of course, is in supposing that heat is causally produced by the increase in mean kinetic energy of the molecules when in fact heat is *constituted* by their mean kinetic energy. The causal reading in this case is simply erroneous. We believe that advocates of interlevel causation often follow Betty in this mistaken line of explanation. The primary virtue of our positive analysis of interlevel relations as mechanistically mediated effects is that it respects the difference between constitution and causation while, at the same time, explaining why it is sometimes tempting, if technically erroneous, to construe the interlevel relationship causally.

## Explicating apparently legitimate interlevel causal claims

Although talk of causation seems problematic when there is a constitutive relation between the putative cause and effect, the mechanistic framework provides an unproblematic way of articulating the idea that gives rise to talk of 'interlevel causation' without appeal to any causal relations that operate across levels. In this section, we first use the example of the eye to illustrate this in both the bottom-up and top-down direction before turning to a number of hypothetical examples intended to reflect the sorts of situation in which appeals to bottom-up and top-down causation are often made.

Begin with the case of bottom-up causation, about which there is a temptation to say that the behavior of the whole mechanism is caused by the activity of its parts. For example, there is a temptation to say that the activation of cyclic GMP phosphodiesterase, which catalyzes the conversion of cyclic GMP to 5′-GMP, causes rod cells to hyperpolarize, which in turn causes the eye to transduce light into neural activity. But the activation of cyclic GMP phosphodiesterase is part of the activity of depolarization, which is part of the eye's transduction of light. However, by attending to the constitutive relationship and mechanistic mediation, we can both trace the intra-level causal processes within the eye and note that the transduction process is constituted by these sub-processes.

Turning to top-down causation, there is likewise a temptation to say that the activity of the parts is caused by the operation of the whole mechanism: when light impacts on the eye and causes it to transduce the visual stimulus into neural activity, it causes the depolarization of the rods. Again, however, by tracking both the constitutive relation between the eye and its parts and tracing the intra-level causal processes within the eye, we can describe what is going on without needing to invoke interlevel causes.

In the remainder of this section, we work through a number of examples in which talk of interlevel causation seems compelling. Each case shows how interlevel causal claims can be translated into less mysterious talk of causation and constitution.

### *Apparent bottom-up causal claims*

In considering possible bottom-up causal relationships, it is necessary to distinguish cases where the seemingly lower-level entity is really part of the mechanism, and hence at a lower level in our sense, from those in which it is not. We start with a case in which it is not, and in which (consequently) a familiar causal analysis is sufficient.

#### *The general's infection*

When a virus kills a general, this seems to be an interlevel causal interaction. On many different senses of 'level,' the virus and the general are at different levels. They are entities of very different sizes, they tend to interact with different sorts of things, and very different forces drive their behavior.

However, the general and the virus are not at different mechanistic levels. The virus is not a component in any of the myriad mechanisms composing the general. This is simply a case where things of different sizes interact. As we noted in Section 'Levels', this is neither uncommon nor problematic. Planets and molecules exert gravitational forces on each other, accounting for the fact that planets have atmospheres. An electrical spark can ignite a tank of gasoline and destroy a building. The general's army can trample and kill a plant as it marches. What view of causation leads one to suppose that intersize causation is somehow problematic? Perhaps one could argue that causes must resemble their effects, and the things at different size scales are too different to interact. Few philosophers endorse this requirement on causation, given that most causes do not resemble their effects one bit (consider matches and fires, the HIV virus and AIDS). But even if the requirement were legitimate, there would need to be a further argument that size is a relevant dimension of difference. Is it perhaps that things of different sizes have different characteristic forces? This seems likely to be true, but one requires a further argument that these forces cannot act (even if only negligibly in some cases) on items at different size scales. Or is it perhaps that all examples of intersize causation, when analyzed, reveal only the smallest parts of things interacting with the smallest parts of things? Even if this were true, it would prove too much in that it involves the rejection of higher-level causes simpliciter, and it therefore has no independent bearing on the question of interlevel causation. Sometimes big things interact with little things. There is no mystery here.

*The general's heart attack*

If instead a heart attack killed the general, the situation would be different. If the general is viewed as a living human being (as opposed to an occupant of an office or rank in the military) whose behaviors are those of ordinary life (locomoting, communicating, etc.), then the heart is a component in one of her bodily mechanisms (e.g., the circulatory system). The failure of blood supply to the heart muscle is a failure of a part of the overall mechanism, and the 'effect' is on the whole mechanism – namely, death. By combining an intralevel conception of cause and a constitutive relation between levels, we can provide an unproblematic account of what transpired: we trace the effects of the infarction through the mechanism. When the heart stops beating, it stops transporting oxygen and nutrients to the other tissues of the body, and they cease to function. We thus explain how a variety of physiological mechanisms cease to function. And insofar as that non-functioning constitutes the general's death, we explain her death. Notice that when we reach the state of the mechanism that constitutes the state of death, we do not say, with Betty Crocker, that it causes death. It just is death. An assassin who charges extra for the general's death – after charging for the cessation of physiological function – is pulling a fast one. Our hybrid explication thus avoids the temptation to attribute causation between a constituent and the mechanism of which it is part.

To return briefly to the virus and the general, there are pragmatic contexts in which a similar interlevel analysis is appropriate in that case. We might, for example, be interested in how the virus killed the general. We are then interested in the mechanism by which the general's overall function is impaired by the virus. To understand this we need to understand that the virus infects the general by lodging in the mucus membrane and diffusing into the body. The virus then takes over the reproductive mechanisms of the host's cells: it penetrates the cell and injects its genetic material into the nucleus. The cell then begins producing copies of the virus, which themselves commandeer other cells for their own purposes. We can trace these changes to various immune responses and consequent symptomologies that ultimately produce the physiological conditions that constitute the general's death.

In these two examples, the putative interlevel causal claim can be fleshed out in a perfectly intelligible way from the 'cause' (the virus or the heart attack) to the 'effect' (the symptoms or the death). Even if there are missing pieces in the analysis, we have no difficulty envisioning how those parts could possibly work. This is why we do not balk at this variety of causal claim: it is a mechanistically mediated effect. The causal claims, when made explicit, are all intra-level. But we can continue to talk about bottom up causal relations when we are being quick or informal as long as we understand that the change at the higher level is mediated by, and explicable in terms of, a mechanism. Nothing mysterious is lurking here.

*Apparent top-down causal claims*

We now turn to putative cases of top-down causation. These too can be analyzed without loss of content as mechanistically mediated effects. The conditions arising at the higher level are constituted by a particular constellation of states of components of the mechanism at the lower level, and their effect is then transmitted in a normal manner through the lower-level mechanism. We start with cases in which a component is contained in a higher-level system or mechanism and is simply carried along for the ride, and we then proceed to cases in which the lower-level constituents are enlisted in the activities of the higher-level mechanism.

*Ignatius and his hotdogs*

Ignatius, with much labor and strain to his valve, coaxed his hotdog cart to the corner. The cart was full of hotdogs. What caused the hotdogs (and the molecules in the hotdogs, and the atoms comprising the molecules, and so on) to arrive at the corner? Ignatius. The hotdogs (and the molecules, etc.) were part of the cart that he labored to bring to the corner, and when the cart arrived, so did the hotdogs (and their molecular constituents, etc.; cf. Kim 2000, p. 312).

In the sense that we normally think about mechanisms, this is a deviant case. The cart is not doing anything in virtue of which the hotdogs move. But if we are curious why the hot dogs moved with the cart (whereas neutrinos that happened to be in the cart would not necessarily move with it) we need to focus on the interaction of the hotdogs with the walls of the cart, and so on. Even though this is a deviant case in which the parts are not working parts, it nonetheless introduces a common way in which constituents are affected by higher-level changes: they are *carried along for the ride*.[12] The hotdogs are merely along for the ride in the cart, just as the virus is along for the ride as the General moves through the troops. This is a perfectly intelligible (and quite common) variety of explanatory relationship. But it is strained to think of the moving of the cart as causing the movement of the hotdogs. Since the hotdogs are simply part of what Ignatius brought to the corner, we need not refer to some mysterious interaction between the motion of the whole and the motion of the parts. The motion of the whole just is the motion of the parts. If you paid Ignatius to move the hotdogs in addition to paying him to move the cart, he made out on the deal.

This relationship between the cart and the hotdogs, while mechanistically explicable, is not explicable in terms of lower-level mechanisms. The work here is not being done by the atoms in the hotdogs, or by the hotdogs, or by the cart. Ignatius (as he will tell you) is doing all the work. Even if the forces between the hotdogs and the cart are part of the explanation for the arrival of the hotdogs, they alone would not get the hotdogs to the corner. They need Ignatius for

[12] Lindley Darden, personal communication, describes genes as being merely along for the ride when the centrosomes pull the chromosomes apart during mitosis.

that. When something is carried along by the object of which it is a part, the explanation of the transport of that object is (excepting the forces between containers and their contents) sufficient, and there is no gain (and considerable loss) in trying to spell out all the causal processes in terms of the parts. There is a perfectly intelligible etiological relationship (Ignatius pushing the cart) and a perfectly intelligible constitutive relationship (the cart's containing the hotdogs, their molecules, etc.). And this is all there is to be said.

As we will see, this same hybrid story also applies to cases in which changes to the mechanism as a whole result in (that is, have mechanistically mediated effects upon) the behavior of the working parts in the mechanism.

*Hal's glucose metabolism*

Hal steps onto the court, serves, and so begins the tennis match. Very quickly, blood borne glucose is taken up through the cell membrane. Once inside, it is phosphorylated and bound into molecules of hexosediphosphate. This is not a case of simply being carried along for the ride. Hal's muscle cells are, it is true, carried along when he swings his racket. But Hal's tennis-playing also alters the behavior of innumerable biochemical pathways and cellular mechanisms that are involved in his tennis playing, both in the short-term and in the long-term. Why did Hal's cells start using more glucose (i.e., binding glucose into molecules of hexosediphosphate)? Because Hal started to play tennis. Similar stories could be told about Hal's respiratory mechanisms, visual system, and many others besides. Changing the behavior of the mechanism as a whole changed the activities of its components. It may be appropriate to say that the components are along for the ride, but if so, this is a different, more active, kind of ride than Ignatius's hotdogs received. Hal's glucoregulatory mechanisms are *enlisted* in the ride.

This is the sort of case for which appeal to top-down causation seems most compelling. However, the case can be described without remainder by appeal only to intra-level causes and to constitutive relations: the 'effect' of the tennis match on glucose metabolism is mediated by a mechanism. In outline: When Hal started to play tennis, the nerve signals to the muscles caused them to metabolize the available ATP to ADP to provide the energy to contract the muscle cells. The increase in ADP made it available as a receptor for phosphates in high-energy bonds in 1,3-diphosphoglycerate produced at the end of the glycolytic process. This allowed a cascade of reactions earlier in the pathway to proceed, eventually allowing a glucose molecule to take up a phosphate from another ATP molecule, initiating the glycolysis of that molecule.

In this and many similar cases, a change in the activity of the mechanism as a whole just is a change in one or more components of the mechanism which then, through ordinary intra-level causation, causes changes in other components of the mechanism. Hal's playing tennis is in part constituted by activities at neuromuscular junctions, and activities at those junctions cause, in a perfectly straightforward etiological sense, changes in the organization and behavior of

cellular mechanisms.[13] Even if this is made possible by the incorporation of complex interactions, including feedback loops, inside the mechanism, there is nothing mysterious about appealing to a change at a higher level to explain a change at a lower level. Once we have described the mechanism mediating the effect, the drive to speak of this as a case of top-down *causation* vanishes, although such language might be useful as shorthand. As long as an analysis like that offered above is available, there is nothing problematic in so using it.

*A boy and his sled*

A boy's mother buys him a sled. When later he is forcibly removed from both his mother and the sled, the event leaves a permanent memory trace. Through his adolescence and adulthood, this memory comes to stand for him as a singular emblem of motherly love and profound loss. At the moment of his death, the now old man recalls a winter afternoon on his sled, utters the word 'Rosebud,' and falls silent. Here again is an apparent case of top-down causation. The young boy's tragic experience leaves a permanent trace in the neural and molecular circuitry of his brain.

Absent from Wells' account (thank goodness) is any hint of Kane's neurophysiology. The interesting features of Kane's memory, and the explanation for its poignancy, lie in its intentional content. Skeptics who believe that the normative or the intentional cannot be explained, without remainder, by detailing causal mechanisms (no matter how complete and complex) will reject an explanation in terms of mechanistically mediated effects in this case, but in so doing, they also reject the idea that the normative or intentional phenomena are at a higher mechanistic level than the neural (or whatever) level below them. We will not try to rebut such skeptics here, but merely note that the issue of top-down causation between levels of mechanisms does not arise for those who reject mechanistic explanation of the normative or intentional (just as the issue does not arise for advocates of strong emergence).

Suppose, however, that such a mechanistic explanation (whatever that would amount to) were possible. And suppose that Kane's episodic memory system works the way that neuroscientists think it does (see, e.g., Eichenbaum 2002). Then Kane's memory storage might go something like this: sensory input at the time of the event resulted in a pattern of activity across different regions of Kane's cortex, activity corresponding to different features of the event in question. Some of these features are rather abstract, such as his feelings toward his mother, his father, and toward the ideas of ownership, power, and control. Some correspond to rather particular and fleeting features of the young Kane's world at the time (the chill of the wind, the sound of a voice, the taste of anger). However this representation is achieved, it is preserved in the neural networks of the hippocampus (by virtue of the influence of cortical input on the strength of synaptic connections in that brain region). Over time, repeated activation of

[13] In fact, one might differentiate a hierarchy of mechanisms in such a case: the match moved the arms, contracted the muscles, metabolized the glucose, and phosphorylated the enzyme.

the hippocampus (perhaps during REM sleep and surely during future cases of remembering) re-excites these cortical patterns. The repeated re-excitation of this basic pattern (certainly shaped and altered with each occurrence) consolidates the representation in the cortex by changing the strengths of connections between cortical cells under this 'tutoring' activity of the hippocampus. So Kane leaves his childhood home for the city, his hippocampus starts to encode representations, the synapses in the hippocampus and the cortex begin to change their strengths, and myriad molecular mechanisms within the cells begin to churn in their own distinctive ways. And why did all of this happen? Because Kane lost his mother and his sled in the same crushing event.

Assuming that our mechanistic story about Kane's memory system is right, his cortex must exhibit a kind of organization that is able to respond to such subtle and abstract features of the human social environment as his relationship to his mother, his father, and his benefactor; to his sudden elevation in social stature; to his institutional affiliations, friendships and marriages; to his need for control and power. Further, his memory system must be able to behave in certain complex ways: to reinforce a memory, to distort it, to associate it with other memories. And last, this memory system must be capable of producing novel kinds of effects on other cognitive systems, such as those responsible for self-monitoring and planning. So Kane's episodic memory mechanism is put into a new environmental situation, one to which the mechanism as a whole is organized to be responsive, and the effects of this change in environment ramify through the other components of the memory system, including the hippocampus, its cells, and their constituent molecules. And just as Hal's glucose metabolism is enlisted in the match, so the hippocampus, the cells and the molecules are enlisted in Kane's ride down memory lane.

### *The general strategy for translating interlevel causes to mechanistically mediated effects*

Each of the above cases of putative interlevel causation is rendered intelligible and reasonable by the same simple strategy. In bottom-up cases, we show that ordinary causal interactions between components of a mechanism produce a condition in the mechanism that constitutes a state of the mechanism identified at the higher level. There are no causal interactions beyond those at a level. In each top-down case, we show that the lower level components are simply being 'carried along for the ride.' In the most trivial cases, the parts are literally carried along for the ride simply by virtue of being moved along with the whole. This goes especially for Ignatius and his hotdogs. In more complicated cases, the parts are enlisted in the ride. As the mechanism as a whole is put into new conditions, it is organized such that its components change with those conditions. Still the same strategy works: In each case, the putative interlevel claim is analyzed into a causal claim coupled with one or more constituency claims. Mechanistically mediated effects, as hybrids of causal and constitutive

claims, can be tested with the four Mill-inspired experimental strategies discussed in Section 'Interlevel causes', and this is compelling evidence that a causal claim is being evaluated, but this is not evidence that the causes at work are, strictly speaking, interlevel. The interlevel relationship is a relationship of constitution. Where there are mechanistically mediated effects, there is no need for the mysterious metaphysics of interlevel causation at all.

## Conclusion

We have proposed a strategy for understanding many interlevel causal claims as mechanistically mediated effects. Mechanistically mediated effects are hybrids of constitutive and causal relations, but the causal relations are exclusively intralevel. Because all of the causal relations are intralevel, there is no need to stretch the notion of causation so that it can accommodate interlevel causal relations. The shroud of mystery surrounding interlevel causation arises from the assumption that the interlevel relation in such cases is both constitutive and causal at once. On our view, the interlevel relationship is only constitutive. This hybrid framework provides a way to understand most, if not all, the cases for which appeal to top-down causes seems compelling. There may be cases that cannot be handled by this account, but if there are, those who invoke the notion of top-down causation for them owe us an account of just what is involved. We suspect that such cases will not involve relations between levels in a mechanism but will employ some other notion of level. In terms of mechanistic levels, then, these cases are not cases of interlevel causation, but ordinary cases of intralevel causation.

Although our explication of interlevel causation in terms of mechanistically mediated effects renders reference to top-down causation unproblematic, it does not show that the phenomenon is unimportant. The biological world, and much of the world besides, is populated by multilevel mechanisms. Talk of interlevel causation is merely a misleading way to talk about an explanatory interlevel relationship that, upon close inspection, does not involve interlevel causes.

## Acknowledgements

We wish to thank Kim Sterelny, Kyle Stanford, and an anonymous referee for comments on earlier drafts. An early version of this paper was delivered in two parts at the 2003 Meeting of the International Society for the History, Philosophy, and Social Studies of Biology in Vienna, Austria.

## References

Alexander S. 1927. Space, Time and Deity, Vol. 2. Macmillan, London.

Anderson P.B., Emmeche C., Finnemann N.O. and Christiansen P.V. (eds) 2000. Downward Causation. Aarhus University Press, Denmark.

Bechtel W. 2006. Discovering Cell Mechanisms: The Creation of Modern Cell Biology. Cambridge University Press, Cambridge.

Bechtel W. (in press). The epistemology of evidence in cognitive neuroscience. In: Skipper R. Jr., Allen C., Ankeny R.A., Craver C.F., Darden L., Mikkelson G. and Richardson R. (eds), Philosophy and the Life Sciences: A Reader, co-editor with responsibility for section on philosophy of medicine, MIT Press.

Bechtel W. and Richardson R.C. 1993. Discovering Complexity: Decomposition and Localization as Scientific Research Strategies. Princeton University Press, Princeton, NJ.

Boogerd F.C., Bruggeman F.J., Richardson R.C., Stephan A. and Westerhoff H.V. 2005. Emergence and its place in nature: a case study of biochemical networks. Synthese 145: 131–164.

Campbell D.T. 1974. 'Downward causation' in hierarchically organized biological systems. In: Ayala F.J. and Dobzhansky T. (eds), Studies in the Philosophy of Biology, Macmillan Press Ltd, NY.

Churchland P.S. 1993. Can neurobiology teach us anything about consciousness? Proc. Addresses Am. Phil. Assoc. 67: 23–40.

Churchland P.S. and Sejnowski T. 1992. The Computational Brain. MIT Press, Cambridge, MA.

Craver C.F. 2001. Role functions, mechanisms, and hierarchy. Phil. Sci. 68: 53–74.

Craver C.F. 2002. Interlevel experiments and multilevel mechanisms in the neuroscience of memory. Phil. Sci. 69: S83–S97.

Craver C.F. Explaining the Brain. Oxford University Press, Oxford (forthcoming).

Dowe P. 2000. Physical Causation. Cambridge University Press, New York.

Eichenbaum H. 2002. The cognitive neuroscience of memory: an introduction. Oxford University Press, Oxford.

Emmeche C., KØppe S. and Stjernfelt S. 2000. Levels, emergence, and three versions of downward causation. In: Anderson P.B., Emmeche C., Finnemann N.O. and Christiansen P.V. (eds), Downward Causation, Aarhus University Press, Denmark, pp. 13–34.

Faye J. 2005. Backwards Causation. Stanford Encyclopedia of Philosophy, http://www.plato.stanford.edu/entries/causation-backwards/ Accessed November, 2005.

Glennan S.S. 1996. Mechanisms and the nature of causation. Erkenntnis 44: 49–71.

Hitchcock C. 2003. Of Humean bondage. Br. J. Phil. Sci. 54: 1–25.

Kandel E., Schwartz E. and Jessel T.M. 1991. Principles of Neuroscience. Elsevier.

Kim J. 2000. Making sense of downward causation. In: Anderson P.B., Emmeche C., Finnemann N.O. and Christiansen P.V. (eds), Downward Causation. Aarhus University Press, Denmark, pp. 305–321.

Lewis D. 2000. Causation as influence. Reprinted in Collins J., Hall N. and Paul L.A. (eds), Causation and Counterfactuals. MIT Press, Bradford.

Machamer P., Darden L. and Craver C. 2000. Thinking about mechanisms. Phil. Sci. 67: 1–25.

Morgan G.L. 1927. Emergent Evolution. Williams & Northgate, London.

Oppenheim P. and Putnam H. 1958. Unity of science as a working hypothesis. In: Feigl H., Scriven M. and Maxwell G. (eds), Concepts, Theories, and the Mind-Body Problem, Minnesota Studies in the Philosophy of Science II, University of Minnesota Press, Minneapolis, pp. 3–36.

Price H. 1996. Time's Arrow and Archimedes' Point. Oxford University Press, Oxford.

Salmon W. 1984. Scientific Explanation and the Causal Structure of the World. Princeton University Press, Princeton.

Reichenbach H. 1958. The Philosophy of Space and Time. Dover, New York.

Sperry R.W. 1976. Mental phenomena as causal determinants in brain functions. In: Globus G., Maxwell G. and Savodnik I. (eds), Consciousness and the Brain: A Scientific and Philosophical Inquiry, Plenum Press, New York, pp. 247–256.

Wimsatt W.C. 1976. Reductionism, levels of organization, and the mind-body problem. In: Globus G., Maxwell G. and Savodnik I. (eds), Consciousness and the Brain: A Scientific and Philosophical Inquiry, Plenum Press, New York, pp. 202–267.

Woodward J. 2003. Making Things Happen. Oxford University Press, Oxford.