To appear in *Levels of Organization in the Biological Sciences (*Brooks, DiFrisco and Wimsatt*, eds.)*

Downward Causation and Levels[1]

Jim Woodward
HPS Pittsburgh

1. Introduction.

This paper is a defense of downward (or top-down) causation and, along with this, a discussion of levels—why we sometimes find it useful to think in terms of this notion and what its limitations might be. The connection between these topics arises in part because words like "downward" and "top-down" suggest a picture according to which the world is organized into "levels" with downward causation involving causes that are at a higher level than their effects. One possible view (see. e.g., Eronen, 2013) is that talk of levels makes no clear sense; if so, whatever might be involved in (what we call) top-down causation, it can't literally be causation from an upper to lower level. Put differently, it might seem that a prior challenge facing anyone who talks of downward causation is to provide an account of levels according to which talk of upper and lower levels makes sense and then evaluate whether there is causation from the former to the latter.

For a number of reasons I'm not going to proceed in this way. Although I think that level notions do legitimate work and thus that we should not try to dispense with them, I also doubt that there is any single, consistent account that captures everything that people have had in mind in talking of levels. My view is that levels talk reflects a number of different considerations that are sometimes mutually reinforcing but also can push us to make very different—indeed inconsistent—judgments in assignments of levels. Privileging just one of these notions is likely to seem arbitrary and in any case will fail to do justice to the variety of motivations that underlie levels talk. It is also true, however, that these different notions are interrelated in various complex ways[2]. My focus in this

[1] Thanks to Sara Green, Bob Batterman and Bill Wimsatt for helpful comments on an earlier version. Green's paper in the present volume as well as Green and Batterman (2017) and Green (2018) provide many additional examples of downward causation and of modeling across levels (or scales) in biology. Batterman's paper in this volume provides a number of illustrations of how talk of levels is tied to claims about scale separation and relative informational autonomy (closely connected to what I call conditional independence) and how, at the same time, in multi-scale modeling it is important to understand how information can be passed across scales. I see this work as complementing my own discussion.

[2] The extent to which different criteria for level assignment lead to largely the same results (or not) is an important question on which I touch only in passing. There is a range of possible positions. One might think that, properly understood, different notions of level or criteria for level assignment produce judgments about levels that largely

paper will be on three (of many possible) ways of thinking about levels, which I believe illustrate these claims:

(1.1) A notion rooted in compositional or part/whole relationships.

(1.2) A notion tied to ideas about independence (including what I will call conditional independence) and along with this, strategies for coarse graining and dimension reduction.

(**1.3**) Closely related to this, a notion based on considerations of computational and epistemic tractability.

As I will try to illustrate, lack of clarity about the relation between these different level notions is one reason why the notion of downward causation has seemed problematic. Conversely, thinking about downward causation provides a very useful point of entry into various ways of thinking about levels.

My discussion is organized as follows: I begin (Section **2**) with some brief remarks about the notions of level that will concern me. Section **3** explores what might be meant by downward causation. Sections **4-5** describe some examples which scientists have found it natural to describe in terms of downward causation. Sections **6-7** consider several objections that philosophers have advanced against the possibility of downward causation. I will argue that these objections are either misguided or do not apply to the examples in question. A crucial part of my argument will be that the putative examples of downward causation on which critics have focused are not, for the most part, what scientists have had in mind in talking of downward causation—the critics' objections do not apply to the examples that I give of downward causation. In particular, one idea to which I will be objecting is a picture according to which top-down causation involves a whole causally affecting its parts. I agree that, at least in many of the cases the critics have discussed, this is incoherent but I also don't think it is what one should understand by downward causation.

Once I have sorted out these issues about downward causation, I will then introduce the notion of conditional independence (Section **8**) and use it to motivate some general remarks about levels. Talk of levels is ubiquitous in science and this raises a number of questions: Most obviously there is the question of what might be meant by such talk. A related question concerns the legitimate function (if any) of such talk. Why do scientists apparently find such talk useful? What work does it do? Does it frequently mislead us, as some critics claim? I address such questions in Sections **9-10.**

---

coincide. I believe this may be Bill Wimsatt's view. At the other extreme one might think that the different criteria lead to results that diverge so much that they render talk of levels useless and misleading. My view is somewhere in the middle between these two possibilities, but closer to Wimsatt's views than those of the complete level skeptics. Thanks to Wimsatt for pushing me on this point.

2. Levels[3]

2.1. Levels as Compositional. One familiar notion of "level" is compositional or mereological: objects or entities at a higher level are "composed" or "constituted by" objects at lower levels in a way that generates a hierarchy. Here "composed" means that the lower level objects are (or at least are thought of as) "parts" of higher-level objects—or at least this is the paradigmatic notion of constitution.[4] Textbooks provide familiar illustrations of this idea. Atoms are composed of protons, neutrons and electrons, molecules are composed of atoms, cells are composed of molecules, multicellular organisms are composed of cells, and so on. This is sometimes described as a "wedding cake" model of levels, since reality is regarded as divided into distinct "layers" based on part-whole relationships. We find this idea (among others) in Putnam and Oppenheim's classic paper (1958) and it often seems to be the preferred conception of levels among metaphysically inclined philosophers. It is this notion of level that (I believe) underlies many philosophical objections to downward causation, since it encourages the idea that this involves causation from a whole to its parts.

2.2. Levels and Independence. Another notion of level is tied to claims about independence where (as I will understand this) it is a matter of relations among *variables*. (More pedantically, it is a relation among what in the world corresponds to variables—e.g., magnitudes such as mass and charge but for brevity I will write "variables" in what follows.) According to this conception, variables *X* and *Y* are at different levels when the behavior of *X* is in some sense "independent" of the values taken by *Y*, so that we can ignore (or largely ignore or ignore in many cases) *X* in constructing a causal explanation of *Y,* appealing instead to other variables at some different level. This notion of level is often tied to considerations having to do with the role of "scales"—spatial, temporal and energetic – in constructing theories and models: sometimes when nature is kind we have "separation" or near separation of scales, so that what happens at one length or energy scale can be understood largely independently of what happens at other scales, and this in turn leads us to think of interactions at one scale as at a different level than interactions at other scales[5].

---

[3] Space precludes detailed discussion of Wimsatt's rich and hugely influential early discussion of levels (e.g. 1994), which includes all of the possibilities I discuss and much more. As will be apparent from his paper in this volume, I share with Wimsatt the view that levels are (sometimes) real features of nature as well as his emphasis on the roles of independence and causal interaction in delimiting levels. I also agree that notions of level can be analytically very useful both in scientific theorizing itself and in philosophical reflection on science.

[4] Some writers, such as Craver and Bechtel also think of properties or "activities" as related by "compositional relations". I regard this as problematic, for reasons described below.

[5] This notion of level as tied to independence (or near or relative independence) is also, I think, the primary notion motivating Simon's notion of near decomposability, discussed in Wimsatt's paper in this volume.

As an illustration consider that, for the purposes of understanding what is going on within the nucleus and phenomena such as radioactive decay, two of the four fundamental forces—the strong and weak nuclear forces—are crucial. These forces are very strong at very short spatial scales. Gravity, another fundamental force, is effectively irrelevant for most purposes in modeling nuclear behavior. On the other hand if we are interested in explaining/understanding chemical behavior—how atoms combine and form molecules and compounds—the strong and weak nuclear forces are effectively irrelevant and yet another force—the electromagnetic force—plays a central explanatory role. In many cases, this separation of levels or scales—the fact that nature permits us to construct theories that explain aspects of nuclear behavior that appeal to factors that are different from those that are required to explain chemical behavior, so that we can do nuclear physics without doing chemistry and vice versa—is crucial for successful science. For one thing, without such separation, constructing models of many phenomena would be computationally intractable. This yields one reason why a notion of level is sometimes important in science.

What do we mean when we say for most explanations of chemical behavior, we don't need to invoke information of nuclear forces? Of course if those forces were sufficiently different, stable nuclei (and atoms) would not exist. So we don't mean that facts about those forces are completely irrelevant to chemical behavior. Rather (I suggest) what we have in mind is something like this: for purposes of chemistry, whatever is relevant about such forces can be represented by the values of a small number of variables, having to do, e.g., with the mass and charge of the nucleus. Conditional on the values of such variables, additional more detailed information about the inner goings on of the nucleus is irrelevant to chemical behavior. So the notion of independence in play here is really a notion of *conditional independence*.

More generally, suppose that some of values of variable *X* are causally relevant to (not causally independent of) the values of variable *Y* but it that it is also the case that there is some variable or set of variables Z constructable from *X* by some coarse-graining operation which we can think of as representing the aggregate impact of the Zs) of much smaller dimensionality than *X* such that given the values of *Z*, additional variation in the values of *X* makes no further difference to the values of *Y*[6]. In such cases I will say that *X* is *conditionally independent* of *Y*, given the values of Z. (As will become clear in Section **8** which spells out this notion in more detail, conditional independence here is not conditional probabilistic independence but should be understood in terms of interventionist counterfactuals.) Such conditional independence allows us to appeal just to *Z* in explaining *Y*. For example, conditional on the value of the temperature of a gas, further variations in the kinetic energy of the individual molecules that are consistent with this temperature make, to a first approximation, no further difference to the pressure of the gas or to the values taken by certain other thermodynamic variables. When, as for these thermodynamic variables, this sort of conditional independence holds, we often find

[6] One of the simplest possibilities for such aggregation is some form of averaging as in the thermodynamic example mentioned immediately below. It is important to understand however that there are much more complicated possibilities, including in particular forms of aggregation that take into account or represent information about spatial or temporal correlations. These are discussed in Batterman's paper in this volume.

it natural to say that those variables are at the "same level" and, moreover, at a level that is "different" from the variables used to characterize the individual molecules that make up the gas.

It is important to understand that this second basis for level talk (which interactions are important and which either can be entirely ignored or subsumed into other, lower dimensional variables) is conceptually quite different from the composition based notion of level or from notions tied directly to size considerations. Whether one object A is part of another B is obviously a distinct question from whether features of A are irrelevant, unconditionally or conditionally, to the behavior of B. Nuclei are "parts" of molecules and nucleons are parts of nuclei but, as noted above, detailed information about nuclear forces between nucleons can be safely ignored in understanding the chemical behavior of molecules. Electrons are also parts of molecules but those aspects of the behavior of electrons that have to do with electromagnetic forces are crucial to understanding chemical behavior.

Of course composition/size based considerations and independence considerations are frequently related—the various components of a cell commonly interact more strongly with one another (are not independent with respect to one another, even given further information about some relatively small number of variables) than they do with the more distant components of other cells (the effects of which may be usefully represented by means of some small number of variables). Nonetheless composition relations and relations of independence/dependence only imperfectly track one another: Entities that are small in size relative to larger entities (or that are roughly the same size as components of those larger entities) can affect those larger entities: viruses and bacteria can contribute to the defeat of armies. Conversely (I shall argue) properties possessed by larger entities can causally affect properties of smaller entities including properties of smaller entities that are components of those larger entities, as when the potential difference across a neuronal membrane affects the behavior of the ion channels that are part of the membrane—see Section **5**.

2.3. Levels and Tractability. A third notion (often tied, however, to the independence- based notion **2.2**. above) ties "level" to considerations having to do with computational and epistemic tractability. Roughly speaking, two sets of variables (or phenomena characterized in terms of those variables) S_1 and S_2 will be at different levels in this sense when there are computational and other sorts of epistemic barriers to modeling or explaining systems that can be characterized by one set of variables (e.g., S_1) and relations among them in terms of variables and relationships from the other set (e.g., S_2). For example, because of such barriers multi-compartment models of fine-grained aspects of neuronal behavior cannot be aggregated to produce tractable models of whole neurons. In this respect the two kinds of models and the relationships to which they appeal are at different "levels"—see Section **9** for additional discussion. A closely related point is that some variables are only well-defined or measurable at certain levels.[7]

[7] A common illustration: the usual thermodynamic notion of temperature of a gas is only applicable to a collection of molecules at equilibrium—it is not well-defined for an individual molecule. Similarly, in connection with the Hodgkin-Huxley model in Section

3. Interventionism and Downward Causation

To talk about causal relationships between "levels" we require an account of causation. I will assume an interventionist account: *X* causes *Y* when there is some possible intervention that changes the value of *X* and, along with this, there is an associated change in the value of *Y* that occurs in a "regular" or "uniform" way. Here "regular" means that in some range of background circumstances, the intervention setting the value of $X = x$ is either followed by the same value of *Y* or the same stable probability distribution for the values of *Y*. When *X* is at a "higher" level than *Y*, and this pattern of dependency obtains between *X* and *Y*, *X* downward causes *Y*. (Obviously in characterizing downward causation in this way, we are not assuming a notion of level according to which variables that are causally related are automatically assumed to be at the same level—instead we are assuming some other notion of level such as a composition-based notion.) This how many scientists who make use of the notion of downward causation understand this notion. For example George Ellis (2016) writes:

> One demonstrates the existence of top-down causation whenever manipulating a higher-level variable can be shown to reliably change lower level variables. (16)

We can further flesh out the idea of a "reliable" or regular change in *Y* under an intervention on *X* in the following way.[8] Assume that the upper-level variable *X* has a number of different lower level "realizers." For example, if *T* is the temperature of a gas, many different possible arrangements of gas molecules, characterized in terms of the values of position and momentum variables for each of the molecules composing the gas, will realize the same value of this temperature.[9] Thus a manipulation of an upper-level variable such as *T* which sets it to some value *t* can have lots of different possible realizations, corresponding to many different possible arrangements of gas molecules. However, for this manipulation to have a reliable (or uniform) effect on some second variable *Y*, we require that all of the different realizations of *t* (or almost all of them—see Sections **8-9**) should have the same uniform effect on *Y*, where again this means that they lead to the same value for *Y* or the same probability distribution for *Y*. In other words, given that *T* is set to the value *t*, it should not matter how that value is realized by the values of the associated lower level variable as far as the effect of $T=t$ on *Y* is concerned—the effect of $T=t$ on *Y* should be in this sense "realization-independent." If this condition is not met, *T* will not count as a cause (top-down or otherwise) of *Y*. This

---

5, the membrane potential is only defined (and measurable) as a feature of the whole membrane.

[8] Ellis (2016) imposes a similar condition.

[9] Realization is thus a relation between the *values* of an upper-level variable and various *values* of a lower-level variable with many different values of the latter mapped into a single value of the former. It is *not* a relation between an upper-level variable and many lower-level variables. Realization is present when values of a lower lever variables are averaged to yield a value for an upper level variable but as noted above averaging is not the only form that realization can take.

condition thus excludes "ambiguous" manipulations of candidate cause variables which have different effects depending on how the cause variable is realized (cf. Spirtes and Scheines 2004). As we shall see in Section **8**, this non-ambiguity requirement is a particular instance of the more general conditional independence requirement mentioned previously.

There are additional conditions that also must be met for downward causation to be present. One particularly important condition is that the putative cause and effect variables must be distinct in the right way—this condition is discussed in Section **6**. Within an interventionist framework it is also important, as Ellis specifies, that the causal relata *X* and *Y* are *variables*[10] where the mark of a variable is that it can assume several different values.[11] Variables include quantities like mass, position, voltage and current but they can also be binary or two-valued. According to interventionism, only variables can stand in causal relationships: we haven't clearly specified what causal relationships we are talking about until we have specified the variables they involve. It is crucial to distinguish variables from things or entities. To anticipate an example discussed below, ion channels and cell membranes are things, not variables and thus cannot literally stand in causal relationships. However, it is typically things or entities that stand in part/whole or compositional relationships. This is one reason why, within an interventionist framework and quite apart from further subtleties about causation, downward or otherwise, talk of wholes causing their parts seems incoherent—wholes and parts are things and hence cannot stand in causal relationships. Variables associated with wholes and parts can sometimes stand in causal relationships but, as we shall see, this needn't involve an objectionable kind of whole/part causation.[12]

4. Downward Causation Exemplified

[10] Recall that this is shorthand for whatever in the world corresponds to variables or values of variables.

[11] For more on variables and values of variables, see Woodward (2015; 2016; forthcoming).

[12] I believe the conditions described are necessary for downward causation but I doubt that they are jointly sufficient. Although I lack space for detailed discussion, there is a natural candidate for an additional condition which is motivated by the fact that it seems possible for a variable to satisfy the conditions above and yet to be highly distributed, non-compact or not simply connected and to not correspond to anything that we could measure or manipulate by upper level measurement and manipulation procedures. The additional condition would require that the upper-level variable not have this character—intuitively, that it exhibit a certain kind of coherence, as when placing a gas in a heat bath has a coherent, coordinated effect on its component molecules. I won't try to explore this idea further, since I don't know how to state it precisely and in any case this additional condition seems to be satisfied in all of the examples of downward causation discussed below.

Here are some putative examples of top-down causation—some drawn from recent books[13] and some from other sources.

4.1) The use of mean field theories in which the combined action of many atoms on a single atom is represented by means of an effective potential V rather than by means of a representation of each individual atom and their interaction. Intuitively, V is at a higher level than the atom on which it acts (Clark and Lancaster 2017).

4. 2) The influence of environmental variables including social relations involving whole animals on gene expression within those animals as when manipulating the position of a monkey within a status hierarchy changes gene expression controlling serotonin levels within individual monkeys. Here position within a social hierarchy is thought of (perhaps on the basis of compositional considerations) at a higher level than gene expression affecting serotonin levels.

4.3) A red hot sword is plunged into cold water and this alters the meso-level structure of the steel in the sword—cracks, dislocations, and grains in the sword. The treatment of the sword—heating and cooling—is at a higher level than these mesoscopic changes[14] and the former downward causes the latter. (Example due to Bob Batterman.)

4.4) Energy cascades. When a fluid is stirred in such a way that it exhibits large scale turbulent motion this motion is gradually transferred to motion at smaller scales– from large scale eddies to much smaller scale eddies. The large-scale motion may be on the scale of many meters, the small-scale motions on the scale of a millimeter where they are eventually dissipated as heat. Viscosity related effects dominate at this smaller scale but are less important at larger scales. (Example due to Mark Wilson.)

5. The Hodgkin- Huxley Model as an Example of Downward Causation.

Each of the previous examples is worth extended discussion but to keep things tractable I will largely focus on just one additional example – the Hodgkin-Huxley (H-H) model of the action potential. This describes the factors causally affecting the overall shape of the action potential within an individual neuron. For reasons of space I will not describe the model in detail but the basic idea is that the neuron can be understood as a parallel circuit consisting of a capacitor which stores charge (the potential V across the neuronal membrane functions as a capacitor), a channel that conducts the sodium current I_{Na}, with an associated time and voltage dependent conductance g_{Na}, a channel that conducts a potassium current I_K with time and voltage dependent conductance g_K, and a leakage current I_l which is assumed to be time and voltage independent. Since the channels are "part" of the cell membrane (they are embedded in it, so that the membrane is, at least on a compositional understanding of levels, at a "higher" level than the channels) and the behavior of the channels, including their conductances, is influenced by (among other factors) the potential difference V across the entire membrane, this looks like a plausible case of top-down causation and indeed it is described as such by, e.g., Denis Noble (2006). For future reference we should also note that according to this model, the potential difference V across the cell membrane is itself causally changed by

[13] Valuable recent discussions of downward causation with many additional examples include Ellis (2016) and Noble (2006).

[14] The heating and cooling affect the whole sword, not just components of it.

the various currents that occur in the ionic channels—as these change over time (and with different time courses), the total current *I* changes (in an apparent case of bottom-up causation) and *V* also changes, with these changes in *V* again changing the ion currents. It is this temporally extended pattern of mutual influence that accounts for the action potential. However, despite this apparent causal cycle, the fact that the ion channels are part of the cell membrane, and what is arguably the presence of downward causation, the HH model looks an intelligible causal representation—indeed one that is generally taken to be correct. How (if at all) can we make sense of this?

6. Downward Causation and Distinctness of the Causal Relata

As noted earlier one objection to downward causation is that this involves wholes acting on their parts. This is thought to be objectionable because causes and effects must be "distinct" and wholes and parts are not sufficiently distinct to stand in causal relationships. (Objections of this sort can be found in Bechtel and Craver 2007; Heil, 2017 and many others.)

What is meant by distinctness in this context? David Lewis' views are representative of a common understanding of this notion:

> [For C to cause E] 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)

As this quotation suggests, the tendency in the philosophical literature has been to try to understand the relevant notion of distinctness in terms of the absence of logical relationships (non-implication) or the absence of part/whole relations (spatial or temporal) and similar considerations, which leads immediately to the conclusion that wholes cannot cause their spatial or temporal parts because of a failure of distinctness. For example, the individual H_20 molecules making up a body of water are parts or constituents of that body and one might object, as Heil (2017) does, to the claim that the position or motion of the whole body causes the position or motion of one of its molecular constituents on the grounds that these relata are not sufficiently distinct to stand in a causal relationship.[15] Similarly, Bechtel and Craver (2007) consider cases in which some temporally extended process is present which has a subprocess as temporal proper part or constituent and object to the claim the former can exert a causal influence on the former. To use their example, because "the change in the conformation of rhodopsin is a stage in the signal transduction pathway in visual perception, the change in

[15] How does Heil's claim fit with the existence of energy cascades described in **4.4**? Heil objects to whole/part causation but, to anticipate my discussion below, this is irrelevant to **4.4** since the transfer of energy from larger to smaller eddies takes time, so that the relation between the latter and the former is not a synchronic part/whole relationship.

conformation cannot be a cause of signal transduction" (p. 552). For similar reasons they object to the claim that a mechanism considered as whole (that is as a collection of parts or constituents standing in ordinary causal relations with each other) can exert downward causation on the parts or constituents of that mechanism.

I agree with these claims (about the absence of downward causation in these examples) but do not think that the complaint of failure of distinctness among causal relata applies to the top-down relationship in the HH model or the other examples described in Section **4**. In other words, the relationship between *V* in the HH model and the behavior of the ion channels is *not* like the relationship between a body of water and its constituent molecules or like the relationship a whole mechanism and its parts. To spell this out, suppose that P is a spatial or temporal part of W and let *X* be some variable that characterizes some feature of W and *Y* some variable that characterizes some feature of P. Then I claim that it is entirely possible for *X* and *Y* to be distinct in a way that allows for *X* to cause *Y* despite the parthood relationship between P and W. Indeed in some such cases, it may make no sense to think of the variable *Y* as a part of *X* or as logically or semantically related to it in a way that precludes causation. For example, the ionic conductances *g* don't seem in any intuitive sense to be "part" of the variable *V*—it is hard to understand what this could possibly mean.[16] More importantly even if one thinks of this relationship in terms of parts and wholes, these variables don't seem to exhibit the kind of failure of distinctness that variables like "saying hello" and "saying hello loudly" do. Similarly, variables describing the mesoscopic structure of the sword in (**4.3**) not are in any obvious sense "part" of the variable that describes how the sword has been heated and cooled.

In saying this, I don't mean to deny that variables can fail to be distinct in ways that preclude their standing in causal relationships—the concern about failures of distinctness is a legitimate worry. My point is rather that the usual ways of trying to characterize the kinds of failures of distinctness that matter for causal relatedness in terms of logical or mereological relations don't work very well. Here is a first pass at a proposal about distinctness that seems natural within an interventionist framework and which I have defended elsewhere (Woodward 2015). The proposal is that variables are appropriately distinct (and thus suitable candidates for standing in a causal relationship as far as distinctness considerations go) when they satisfy a condition of independent fixability (**IF**):

> (**IF**) A set of variables **V** satisfies independent fixability of values if and only if for each value it is possible for a variable to take individually, it is "possible" to set the variable to that value via an intervention, concurrently with each of the other variables in **V** also being set to any of its individually possible values by

[16] Craver does attempt to elucidate what it is for one "activity" to be "constitutively relevant" to another by appealing to a "mutual manipulability" criterion. I lack space for discussion but notice that the relationship between *V* and the channel conductances appears to satisfy Craver's criterion, which implies (in my view mistakenly) that the relationship between them is constitutive rather than causal. Moreover, systems with causal cycles appear to satisfy Craver's criterion even though they involve causal relationships.

> independent interventions. Here "possible" includes settings of values of variables that are possible in terms of the assumed, logical, mathematical, or semantic relations among the variables as well as certain structural or space-state relationships.

Thus "possible" in **IF** should not be understood as restricted to combinations of settings that are causally possible, although of course if settings are causally co-possible they are possible *tout court*. For example, in the usual state-space formulation of mechanics, the three-dimensional position and momentum components of each of a collection of particles at a time as well as the components for the same particle at different times are regarded as independently fixable, even though, once dynamical considerations are introduced, certain combinations of these may be causally excluded. Obviously **IF** draws upon (and does not explicate) some antecedently understood notion of possibility that is broader than causal possibility. What **IF** adds is a focus on whether *operations* involving setting of variables to values are co-possible—this turns out to be crucial.

To further illustrate (**IF**), consider Lewis' example of saying "hello" and saying "hello" loudly. Expressed in terms of variables, suppose that X has two values (x_1= saying hello loudly, x_2 = doing something other than saying hello loudly) and Y has values y_1 = saying hello, y_2 = not saying hello.) Then the values x_1 and y_2 are not co-possible and there is a failure of distinctness. Notice that using (**IF**) to reach this conclusion does not require the assumption that saying hello is a "part" or a "constituent" of saying hello loudly (whatever that might mean) although it does require a judgment that it is in the relevant sense not possible (presumably because of logical or semantic relationships) to say hello loudly without saying hello[17]. As another illustration, position and momentum for an individual particle satisfy independent fixability (and hence are distinct), even though some may find it tempting to argue that there is a logical or part/whole relation between these variables (since momentum is the product of mass and the time derivative of position). As this last example illustrates, (**IF**) does not depend on our being able to make sense of constitutive relations among variables and does not always yield the same conclusions as this last notion.

---

[17] Older readers may remember "logical connection" arguments that claimed to show that desires and beliefs were "logically connected" to associated actions and hence not sufficiently distinct to serve as causes of them—e.g., the desire D to drink beer could not cause drinking beer B because of a "logical connection" between the two. A consensus eventually emerged that this was a flawed argument—despite the alleged logical connection between D and B, D can cause B. I see this as illustrating the dangers of relying on unclear ideas about logical connection and overlap in trying to elucidate distinctness among variables. Note that **IF** yields the correct judgment about this case: all of the different values of D and B are compossible: one can have the desire to drink beer without drinking, one can drink without the desire (e.g., out of a feeling of social obligation) and so on. Of course if it is claimed that whenever one drinks, it follows a priori one has the desire (i.e. that drinking without the desire is excluded on conceptual grounds) , (**IF**) will yield the conclusion that the relation is not causal, but this seems the correct assessment.

If we apply (**IF**) to the HH model, the question we should ask is whether *V*, the putative top down cause, and the channel conductances and ionic currents, the putative effects of *V*, are distinct in a way that satisfies (**IF**). The answer to this question is "yes." First *V* is clearly manipulable in a way that is independent of the values taken by the conductances or ionic currents. This does not require questionable judgments about non-causal forms of possibility: it is shown by some of the experiments that were used by Hodgkin and Huxley to establish their model. These involved the use of a newly invented device called a "voltage camp." This enabled the experimenters to impose a stable potential difference (at various levels they were able to choose) across the cell membrane in a way that depended only on the value set by the clamp. The clamp thus functioned as an (arrow-breaking) intervention device, with the membrane potential difference fixed by the device rather than by such endogenous causes as the operation of the ion channels. This allowed the experimenters to see and investigate (isolate) the effect of *V* on the ionic currents and conductances in a way that confirmed the predictions of the HH model. Similarly, various molecular interventions are possible that alter the individual ionic channel currents and conductances independently of *V* when the clamp device is used and these again show behavior in accord with the HH model.

I claim that this independent manipulability, as captured by (**IF**) suffices to show that it is legitimate to think of *V*, and the channel currents and conductances as sufficiently distinct to stand in causal relationships. A similar analysis applies to the other examples in Section **4**. More generally, the fact that claims of top-down causation often involve claims that variables that are predicated of wholes causally affect variables that are predicted of parts of those wholes is consistent with those variables being sufficiently distinct in the sense of **IF** to stand in causal relationships.

7. Causal Cycles

I noted above that the HH model appears to involve a causal cycle, at least if we confine ourselves to the variables employed in the model. A similar observation holds for many other putative examples of downward causation; often (not always) when an upper-level variable *U* is claimed to act on lower-level variable *L*, the value of *U* (perhaps at some later time) will result (causally) from the action of lower-level variables. For example, the position of a monkey in a dominance hierarchy causally affects the animal's serotonin level but that level in turn affects position in the hierarchy—something that can be demonstrated by exogenously increasing an animal's serotonin level pharmaceutically with the result that the animal rises in the hierarchy.

One possible response to worries about cyclicity is to say that "underlying" any cyclic graph is a model with time-indexed variables with temporal lags that is acyclic. If these temporal lags (or the difference between the values of X_t and X_{t+1}) do not matter to the effects we are trying to capture then the use of a cyclical representation may be unproblematic. For example, in the HH model, the response of the ion channel conductances to a change in voltage across the cell membrane (or more accurately, to a change in the membrane at some distance from the channels) is not instantaneous, although this fact is not represented in the HH model. It is plausible that this temporal delay makes no difference to the generic shape of the action potential which is why the model is successful despite omitting such information.

The issues around how to interpret graphs with cycles are complex, and I don't claim that time-indexing is always a satisfactory treatment. (Rather different cases require different treatments, and there are a number of subtleties that I lack space to discuss.) It is worth noting, however, that graphs with cycles sometimes have a straightforward interventionist interpretation along the "usual arrow breaking" lines. That is, one way of interpreting a bi-directional graph

(7.1) $U \leftrightarrow L$

is as follows: (i) if we were to intervene on U, this would break the arrow directed into U from L while preserving the arrow directed out of U into L, thus replacing (7.1) with the following structure.

(7.2) $U \rightarrow L$

If this interpretation correctly describes the causal facts, one would expect L to change as indicated under this intervention on U. Moreover, if one were to intervene on L, this would break the arrow from U directed into L, while preserving the arrow from L into U so that under this intervention (7.2) is replaced with

(7.3) $L \rightarrow U$

Again if (7.3) is correct, U should change under this intervention on L.

In fact, as already noted, this is essentially what was done as part of the experimental confirmation of the HH model. The use of the voltage clamp constitutes an arrow breaking intervention on V and one looks to see whether under such an intervention on V the channel conductances and currents respond in the way described by the equations, which they in fact do. Similarly, interventions on the channel conductances and currents followed by measurement of V can establish the existence of a causal relationship running upwards from these to V. Thinking of these results as implied by (7.1) thus provides a coherent interpretation of that graph. In the case of the relationship between status and serotonin levels we can take a monkey with currently low status and move him to another less competitive troop where because of his abilities he will rise to a higher status (suppose he is bigger than all of the monkeys in the second troop and smaller than many in the first troop) and observe the predicted increase in serotonin levels which provides evidence for top down $U \rightarrow L$ causation. As noted above we can also change his serotonin levels by an exogenous pharmaceutical intervention and observe the resulting change in his status. For any given monkey, his serotonin level and his status in the absence of such interventions or after the effects of the interventions have been allowed to equilibrate will presumably reflect the joint operation of processes operating in both causal directions in which case the bi-directional graph may be particularly appropriate.

It is sometimes claimed that the cyclic graphs are inconsistent with the directionality or asymmetry of causal relationships. It seems to me that this conflates two issues. It is plausible that causal claims have a kind of directionality built into them and this mandates the use of directed (rather than undirected) graphs to represent such

relationships. However a graph can be directed while still containing cycles.

8. Conditional Independence

In Section **2** I briefly introduced a notion of conditional independence. In this section I spell this notion out in more detail and relate it to my previous discussion, explaining how it bears on notions of level and downwards causation.

I begin with an "ideal" case.[18] Suppose that we have a set of variables L_i with very high dimensionality that are causally relevant (by the standard interventionist criterion of relevance) to some explanandum E (or set of explananda E_i), which may be either upper or lower level. Suppose also there is a set of upper level variables U_k of much smaller dimensionality with the following property: interventions on the values of U_k are also causally relevant to the E and, furthermore, conditional on the values of the U_k when these are fixed by interventions, further variations in the values of the L_i, produced by independent interventions, make no difference to (are irrelevant to) the values of E. In a bit more detail, let us say that the variables L_i are *unconditionally relevant* (alternatively, irrelevant or independent) to E if there are some (no) changes in the values of each L_i when produced by interventions that are associated with changes in E. A set of variables L_i is irrelevant to variable E *conditional* on additional variables U_k if the L_i are unconditionally relevant to E, the U_k are unconditionally relevant to E, *and* conditional on the values of U_k when these are fixed by interventions, changes in the value of L_i produced by interventions and consistent with these values for U_k are irrelevant to E.[19] If it is possible to find such a set of variables U_k (and perhaps also if they meet certain additional conditions of the sort gestured at in footnote 7) we can replace the L_i with them insofar as we are just interested in describing difference-making relationships bearing on E—that is, identifying those variables' variations which make a difference for E. The U_k do just as good a job as the L_i in this respect. And of course identifying such difference-making relationships is what explanation and causal analysis is all about according to the interventionist.

Examples that look roughly like this are quite common. To return to an example mentioned briefly in Section **2**, suppose we are interested in explaining the macroscopic behavior of a gas as characterized by such variables as temperature, pressure and volume A given temperature t for the gas will correspond to or can be realized by any one of a very large number of collections of molecules with different positions and momenta—six such variables for each molecule in the gas, so that this variable has over 10^{24} independent dimensions. But (except for a measure zero set of cases) the impact of any of these profiles on the macroscopic variables depends entirely on their aggregate or average behavior which is summarized by the values of the thermodynamic variables. Given the values of the macroscopic variables further details having to do with the exact positions

---

[18] For a closely related set of ideas, see Chalupka et al. (2017). What follows has been substantially influenced by this paper and by discussion with Frederick Eberhardt.

[19] In other words, we are to imagine that the value of $U_k = u$ is fixed by an intervention, while the value of L_i is set via interventions to any value which is consistent with u. If conditional independence holds, these further variations in L_i should have no influence on E.

and momenta of the individual molecules are conditionally irrelevant to many aspects of the behavior of the gas.

Similarly, suppose, as the HH model in effect claims, that as an empirical matter, given the value of the overall membrane potential *V*, further lower-level detail captured by lower-level variables (e.g., variables describing the fields associated with the individual atoms and molecules making up the membrane) is conditionally irrelevant to the shape of the action potential, the gating behavior of the ion channels and so on.[20] To the extent this is true, we may legitimately appeal just to *V* to explain these explananda. Under these conditions, *V* is a legitimate downward cause.

9. Levels and Conditional Independence.

How does all this relate to assignments of levels? In the case of the gas we have a rationale for treating the thermodynamic variables as at the "same" level, since the right sort of conditional independence relation holds among them and separates them from "lower level" information about the position and momenta of individual molecules. In addition, compositional considerations reinforce this assignment of levels. In contrast, in the case of the HH model, compositional and perhaps other considerations suggest that the causal relata (*V* and the channel conductances) are at different "levels," even though there is interaction between these levels. But the underlying logic is the same: it is legitimate to treat (in the right sort of set up) temperature as a cause of pressure because a given value of temperature as uniform effect on other thermodynamic variables like pressure (uniform in the sense that given that value, further variations in molecular arrangements realizing the temperature make no difference to those thermodynamic variables) and it is legitimate to treat *V* rather than some more detailed description of the potential differences resulting from the exact arrangement of electrons along the cell membrane as a cause of aspects of the behavior of the ion channels because the different realizations of *V* have a similarly uniform effect on that behavior.

Several additional remarks may help to clarify how the conditional independence idea is to be understood. Note first that it is relativized to a particular target explanandum (or perhaps a set of these). Variables *L* may be independent of explanandum *E* conditional on the values of variables *U* but *L* may not be independent of some other explanandum *E** conditional on *U*, so that if we wish to account for *E** we do need to take the values of *L* into account. For example, if we wish to account for facts about the specific heats of gas, we must advert to quantum mechanical considerations rather than to macroscopic variables like pressure and temperature. Similarly, if our target explanandum is the overall shape of action potential, then conditional on the variables employed in the HH model, further information about molecular details may be irrelevant but if we wish to explain other features of the system such as the behavior of dendritic trees, this particular conditional independence relation will no longer hold (cf. Herz 2006).

In practice, as some of the examples already discussed suggest, there are often natural groupings of variables for which the same conditional independence relations

[20] Note what this claim says. It does *not* say that there are no local variations in the potential; it says that they do not matter for the explananda of interest.

hold—e.g., conditional independence of various explananda characterized in terms of thermodynamic variables from lower level molecular detail holds conditional on other thermodynamic variables, so that these variables form a natural grouping. This is one basis on which we group whole sets of variables into "levels."

Second, note the form taken by the conditional independence justification of the use of upper-level variables. It is common in the philosophical literature for defenses of upper-level causal claims or explanations to attempt to show that such explanations are *superior* to explanations in terms of lower-level variables and, moreover, superior in a way that is completely independent of "pragmatic" considerations having to do with human epistemic and calculational limitations.[21] The conditional independence justification does *not* claim this. Rather what it attempts to do is to identify conditions under which it is *permissible* or *legitimate* to employ upper-level variables—permissible in the sense that this can be done without explanatory loss. There is no claim that an explanation in terms of lower-level variables (if we could produce one, which is frequently not the case[22]) would be inferior to an explanation in terms of upper-level variables.

Third, the condition described above, involving complete irrelevance of the lower level variables to certain explananda conditional on the values of the upper level variables, is obviously a kind of limiting case, although it is arguably not as rare as some philosophers suppose. The requirement of complete irrelevance may be relaxed in various ways.[23] We might require instead that the L_i be irrelevant to E conditional on U_k for most or "almost all" values of these variables or for values of those variables that are most likely to occur (perhaps around here right now). We might require that for those values of L_i for which exact conditional irrelevance fails, near conditional irrelevance or independence holds—most, even if not all, of the variance in E explained by L_i is explained by U_k. We might think in terms of conditional irrelevance holding on some appropriate time or spatial or energy scale even if not on others—for example, perhaps there are very fast variations in L_i occurring on a very fine-grained temporal scale that can make a difference to E even conditional on the values of the upper-level variables but the L_i very quickly settle down to constant equilibrium values which have an upper level representation for which conditional irrelevance holds.[24]

My argument so far has been that considerations about conditional independence can be invoked to explain why it is permissible or legitimate to formulate causal claims in

---

[21] For recent claims of this sort see Weslake, 2010 and Franklin-Hall, 2016.

[22] See my discussion below.

[23] Of course, to the extent that we do this, we allow for manipulations that are in some respects ambiguous, so there is a trade-off around these considerations.

[24] It is an important general fact about independence and conditional independence relations that they can be scale or grain relative in the sense that switching to different temporal or length scales or adopting certain procedures for aggregating lower-level variables can replace situations in which variables are dependent with situations in which related upper–level variables are independent or conditionally independent. For example, X might be correlated with Y on a very long time scale but if relatively short time scales are what are relevant to the behavior of interest it may be appropriate to treat X as constant, in which case it will be independent of Y.

terms of upper-level variables, including causal claims that involve lower-level variables as effects—we may lose little or nothing by doing so in terms of the identification of difference-making factors for the effects in question. Moreover, such considerations provide one important basis for grouping variables into levels and for understanding when it is legitimate to collapse lower-level variables into more coarse-grained upper-level variables with fewer degrees of freedom. However, of course there is more to the story about why we actually employ such upper-level variables. It is at this point that various sorts of limitations of us humans (and perhaps all bounded agents) come into the picture. Some of these are calculational—we can't solve the 10^{23} body problem of calculating bottom up from the behavior of individual molecules to the aggregate behavior of the gas. In addition, we face the epistemic problem that we are unable to make the kinds of fine-grained measurements that would be required for such calculations to reach reliable results[25].

To take another example, although there are fine-grained neural models employing large numbers (up to 1000) of individual "compartments" (each of which represents a distinct circuit structure for a small portion of the neuron) that can be used to account for aspects of dendritic behavior and the role of neuronal spatial structure, these multi-compartment models cannot, for reasons of computational tractability, be "aggregated up" to produce a model of the whole neuron. For that we require a different model like the HH model which is a "single compartment" model that neglects the spatial structure of the neuron but nonetheless is adequate to explain the overall shape of the action potential.[26] We thus find that not only is it permissible to formulate theories in terms of upper-level variables if we wish explain certain explananda but that we have often have no choice but to do this if we want models that are tractable or that we can

---

[25] Wimsatt makes the important point in correspondence that in the biological realm such computational and epistemic limitations influence what are the correct causal relations and not just how we model these. If an organism can only perceptually detect relatively coarse-grained differences in, say, a prey or predator because of such limitations, then it is these coarse-grained features that causally affect behavior, rather than some finer-grained variable. My flight behavior is causally sensitive just to whether the animal before me is a tiger, rather than to fine grained details of molecular realization.

[26] Cf. Herz et al (2006):

> Single-compartment models such as the classic Hodgkin-Huxley model neglect the neuron's spatial structure and focus entirely on how its various ionic currents contribute to subthreshold behavior and spike generation. These models have led to a quantitative understanding of many dynamical phenomena including phasic spiking, bursting, and spike-frequency adaptation.

This short paper is very interesting in its discussion of neuronal models at different "levels" and the way in which models at each level are able to capture some aspects of neuronal behavior and not others and how "abstraction" (which basically amounts to neglect of certain features of the neuron which are irrelevant to the behavior one is trying to understand) can lead to models that can account for aspects of higher level neuronal behavior that, for computational reasons, cannot be captured by lower level models.

calculate with. Put differently, we are sometimes in the fortunate situation that nature presents us with relations of conditional irrelevance/independence that we can then exploit to construct tractable models which would not otherwise be possible. When we build models and theories that exploit these opportunities, they will be structures in which upper-level causation appears. Note that although such computational considerations may reflect, at least in part, facts about us, the facts about conditional independence or near approximations to it which they exploit have to do with what nature is like—the latter are not just reflections of our computational limitations.

Considerations having to do with calculational and epistemic constraints of the sort just described thus represent another set of considerations (briefly described under **2.3** above) that influences judgments about levels. Among other considerations, models at different levels of detail may employ very different varieties of mathematical description that are difficult if not impossible to stich together smoothly with the consequence that we cannot straightforwardly extend models and theories that are successful in accounting for the behavior of systems at certain scales (or levels of detail) to behavior at other scales or levels of detail. For example, models at one level may employ partial differential equations (which may be used to capture the role of spatial structure), at another level ordinary differential equations (which may abstract away from spatial structure), at another level Boolean or structural equations (which will neglect the underlying dynamics described by the differential equations), and at still another level Bayesian representations in terms of probability theory.[27] Thus we often end up with situations in which each of a variety of different kinds of models have their own distinctive explananda which they account for and other possible explananda which they cannot explain, either because the right sorts of conditional independence relations do not hold or for computational reasons or both. Again, this encourages us to think of such situations in terms of a separation into levels. The problem we then face is getting these levels to "talk to one another" when conditional independence partially fails[28].

To relate these ideas to my earlier discussion of downward causation, consider the question of why we have a notion of downward causation at all and regard claims of downward causation as sometimes legitimate rather than (as some skeptics claim we should) insisting that the only true literally causal claims are those that relate variables that are all at the same lower level. My answer is that (1) when the candidate top-down cause has a uniform effect on some other variable regardless of how it is realized, we lose nothing by describing the situation in terms of upper-level rather than lower-level variables and (2) for computational reasons, we may not be able to formulate an account of the effect in terms of lower level variables in any case.

Finally a brief remark about "autonomy". This word is used in many ways but one natural meaning is that a framework or theory is autonomous or relatively so to the extent that one doesn't need information coming from some other theory or level to adequately model some range of phenomena. For example, as Batterman observes in his paper for

---

[27] Thus in Herz et al.'s (2006) catalog of different kinds of models, their level 1-3 models employ various kinds of differential equations while their level 5 models black box entire neurons and treat them as computing via Bayesian updating.

[28] Some strategies for dealing with this problem are discussed in detail in Batterman's paper in this volume.

this volume, the Navier- Stokes equations are autonomous with respect to many explananda concerning fluid behavior in the sense that they account for those explananda without requiring information about the molecular details of the fluid. Obviously autonomy, when so understood is a relative or conditional notion in several senses. A theory may be autonomous in its ability to account for one set of explananda but not others, as Batterman also observes. A theory may need a whole lot of information from some other source to be adequate or it may need relatively little information, perhaps of a very non-detailed generic sort. To pick up on the quotation at the beginning of Sara Green's paper, certain generic facts about cars and the behavior of their drivers may be relevant to modeling traffic flow, but not the details of the working of internal combustion engine. I think of my remarks about conditional independence as one possible way of capturing these ideas[29].

10. Conclusion

My remarks in the previous section attempt to provide a partial answer to the questions about the function of level talk in science posed in Section **1**. In addition to the role played by compositional considerations, such talk can be motivated by empirical facts about conditional independence and by considerations having to do with what it is

[29] I cannot resist two further remarks. First, ideas related to conditional independence might be used to capture part of what may be meant by talk of "emergence" and in a way that renders that notion unmysterious. That certain lower level information is conditionally irrelevant to certain explananda should not be metaphysically puzzling and does not by itself imply the explananda are inexplicable in terms of the lower level information. Second, traditionally discussions of autonomy have been closely bound up with issues about reduction. Fodor, for example, says that psychology is autonomous to the extent that it is not reducible to neurobiology, where by "reduction" he has in mind something like Nagelian reduction. Understanding autonomy in terms of conditional independence does not map onto Fodor's picture in any simple way. If a psychological theory is fully type reducible to a true neurobiological theory, this would presumably mean that the psychological theory by itself was fully adequate in accounting for the psychological phenomena it was meant to explain since all the generalizations of the psychological theory follow from the true neurobiological theory. In this case the psychological theory would be autonomous in the sense I propose. Similarly, although multiple realizability is often taken to undermine the possibility of Nagelian reduction, it is compatible with autonomy in the sense described above. (Multiple realizability is also compatible with failure of autonomy if realization independence fails.) If on the other hand, the psychological theory was empirically inadequate by itself and required extensive supplementation or correction by neurobiological information in order to be adequate, the psychological theory would not be reducible to the neurobiological theory but it wouldn't be autonomous either. What this shows is that the extent to which the relevant information in a lower level theory is captured by the categories in an upper level theory(so that additional input from the lower level theory is not required) is very different from whatever is captured by Nagelian reduction.

possible to represent and calculate using various sorts of mathematical models. I see these factors as working together and interacting—the conditional independence facts create niches or opportunities for computationally tractable models which succeed in explaining certain effects in virtue of abstracting away from certain conditionally irrelevant factors at a more fine grained levels of analysis.[30] In addition, ideas about levels can also play the useful heuristic role of providing plausibility arguments to theorists or modelers about which factors they may be able to ignore, prior to the construction of detailed models although of course such arguments always need to be checked empirically. These are all considerations that help to explain why it is sometimes justifiable and indeed salutary and advantageous to make use of level-based arguments and reasoning strategies.

That said, we should also recognize the following complicating (and sometimes countervailing) considerations which suggest caution about too much reliance on level-based considerations: First, finding cases in which conditional independence holds even approximately is (at least typically) not easy—it requires finding the "right" variables and the right strategies for representing the impact of high dimensional variables in lower dimensional ways. In some domains of inquiry, it may not be possible to find such variables at all (or at least variables that are well-behaved in the sense of being cognizable and measurable) —instead many different variables which we think of as at very different levels (where this is assessed in terms of compositional or other considerations) may all matter to the effects we are trying to explain, so that there is extensive "causal leakage"[31] across levels.[32] This does not make explanation impossible but it certainly makes it more difficult. In such cases, heuristics based in ideas about sharp separations of levels can mislead.

We should also bear in mind that, as urged earlier, conditional independence facts are explanandum relative—from the fact we can legitimately neglect certain factors in accounting for certain explananda *E*, it does not follow that we can legitimately neglect those factors in accounting for some other explanandum *E**, even if *E** seems, intuitively similar to *E* or at the same level according to some notion of level such as one based on compositional considerations. Whether we can legitimately neglect such factors in accounting for *E** is always an empirical issue, which cannot be settled a priori. We should be particularly sensitive to the possibility that conflations among different notions of level can lead us to assume conditional independence in cases in which it is not warranted, as when we assume that composition-based differences in level automatically warrant assumptions about conditional independence.

References

---

[30] This notion of computational opportunities arising from possibilities of avoiding modeling in detail various aspects of the systems with which we deal is developed in very rich detail in Wilson, 2017.

[31] See Wimsatt (1994) who attributes the phrase to Stuart Glennan.

[32] As an illustration consider the model of the causation of major depression in men developed in Kendler et al. (2006). This employs variables spanning many different levels: genetic risk, personality variables such as low self esteem, conduct related variable such as substance abuse and social or environmental variables having to do with e.g. early parental loss. None of these "screen off" the effects of the others on depression.

Bechtel, W and Craver, C (2007) "Top-down Causation Without Top-Down Causes" *Biology and Philosophy* 22:547–563.

Chalupka, K. Eberhardt, F., and Perona, P. (2017) "Causal Feature Learning: an Overview" *Behaviormetrika* 44:137–164.

Clark, S. and Lancaster, T. (2017) "The Use of Downward Causation in Condensed Matter Physics" In M. Paleotti and F. Orilia (eds.) *Philosophical and Scientific Perspectives on Downward Causation*. New York: Routledge.

Ellis, G. (2016) *How Can Physics Underlie the Mind? Top-Down Causation in the Human Context*. Berlin: Springer.

Eronen, M. I. (2013). "No Levels, No Problems: Downward Causation in Neuroscience" *Philosophy of Science* 80(5), 1042-1052.

Franklin-Hall, L. (2016) "High Level Explanation and the Interventionist's 'Variables Problem'" *British Journal for the Philosophy of Science* 67:553-577.

Green, S. (2018) "Scale Dependency and Downward Causation in Biology" *Philosophy of Science* 85

Green, S. and Batterman. R. (2017)"Biology Meets Physics: Reductionism and Multi-Scale Modeling of Morphogenesis." *Studies in History and Philosophy of the Biological and Biomedical Sciences* 61:20–34.

Heil, J. (2017) " Downward Causation" In Paoletti, M. and Orilia , F (eds.) *Philosophical and Scientific Perspectives on Downward Causation*. New York: Routledge, 42-53.

Herz, A., Gollisch, Y., Machens, C. and Jaeger, D. (2006): "Modeling Single-Neuron Dynamics and Computations: A Balance of Detail and Abstraction", Science, 314: 80–5.

Lewis D. (2000). "Causation as influence". Reprinted in Collins J., Hall N. and Paul L. (eds), *Causation and Counterfactuals*. Cambridge: MIT Press

Noble, D. (2006) *The Music of Life.* Oxford: Oxford University Press.

Oppenheim, P., and Putnam, H. (1958). "The Unity of Science as a Working Hypothesis" . In Feigl, H., Scriven, M. and Maxwell, G. (Eds.). *Concepts, theories, and the Mind-Body Problem*. Minneapolis: University of Minnesota Press, 3-36.

Weslake, B. (2010) "Explanatory Depth" *Philosophy of Science* 77:273-294.

Wilson, M. (2017) *Physics Avoidance: Essays in Conceptual Strategy*. Oxford: Oxford University Press.

Wimsatt, W. (1994) "The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets", in *Biology and Society: Reflections on Methodology*, M. Matthen and R. Ware, *Canadian Journal of Philosophy*, supplementary volume 20: 207–274.

Woodward, J. (2015) "Interventionism and Causal Exclusion" *Philosophy and Phenomenological Research* 91: 303- 347.

Woodward, J. (2016) "The Problem of Variable Choice" *Synthese* 193:1047-1072

Woodward, J. (Forthcoming) "Explanatory Autonomy: The Role of Proportionality, Stability, and Conditional Irrelevance". *Synthese*