Preprint - Accepted for presentation at the 29th Philosophy of Science Association Conference
New Orleans, LA

The Philosopher's Error: Levels, Causation, and Constraints

Ramy Amin[1]

Word count: 3936

Abstract

Downward causation plays a central role in the debate around levels of mechanism. Both levels' enthusiasts and skeptics reject it, arguing that it is incoherent to conceive of wholes causing the parts which constitute them. In this paper, I advance an argument from causal constraints against claims of the unintelligibility of constitutive downward causation, arguing that constitution relations neither exhaust the totality of relations that a proper whole is subject to, nor do they preclude another type of relation that a proper whole can have with respect to another proper whole.

1. Introduction

Claims of the unintelligibility of interlevel causation are a central talking point for levels of mechanism enthusiasts. Craver (2014) states that "mechanistic levels are not causally related to one another" (21). Craver and Bechtel (2007) argue that talk of interlevel causation is unintelligible and provide a mechanistic alternative in which bottom-up causation is spelled out in terms of constitution relations, and top-down causation is translated into "mechanistically mediated effects" (548). In particular, the latter claim regarding top-down causation has been taken for granted by both level enthusiasts and skeptics. Eronen (2021), who is a levels skeptic, states that Craver and Bechtel's "reasoning against downward causation does not carry over to contexts where downward causation is understood in a noncompositional way" (83), therefore he accepts downward causation only in instances where the two relata are not compositionally related. Even Woodward (2021a), who

[1] Department of Philosophy, University of Cincinnati, Cincinnati, OH, USA.

explicitly argues for top-down causation, still joins Craver, Bechtel and Eronen in rejecting "a picture according to which top-down causation involves a whole causally affecting its parts", citing that this is incoherent, and that this is not how one should understand top-down causation (176).

By and large, levels enthusiast and skeptics seem to agree on the unintelligibility of constitutive downward causation. Two main worries are behind this agreement. The first is that it is absurd to talk of wholes causing their parts: if parts constitute a whole, then the relationship between the two is one of *realization* rather than causation. As Craver (2014) explains, "[t]he beating of the heart is realized, not caused, by the choreographed movements of the auricles and ventricles. It is awkward and unnatural to assert otherwise" (11). The second is a worry about strong (or what Craver calls spooky) *emergence*. The whole *is* its parts, and to say that the whole can stand in a causal relation to its parts means that the whole has a unique property that is above and beyond the contribution of the parts. This is taken to be a commitment to "the existence of higher-level properties that have no explanation in terms of the parts, activities, and organizational features of the system" (Craver, 2014, 21).

Despite the agreement between levels interlocutors on the unintelligibility of constitutive downward causation, my goal in this paper is to challenge such agreement. The idea here is that if one of the essential properties of levels of mechanism is that they are causally unrelated, yet we can show instances of constitutive downward causation, then that puts pressure on the concept of levels of mechanism, and forces mechanists to confront the fact that they either need to adjust their claim, allowing at least for some interlevel causation, or find new ways to argue for levels of mechanism without relying on claims about the unintelligibility of interlevel causation.

I first unpack the notion of mechanistic constitution, which I take to be the main reason behind the insistence on denouncing interlevel causation. I then present Ross's (2023) concept of causal constraints and argue that it challenges the rejection of constitutive downward causation. Finally, I reflect on an example of a causal constraint from neuroscience where the brain's structural connectome, which stands at a higher-level than the mechanisms constituting it, exerts influence over the operations of lower-level

mechanisms in an instance of constitutive downward causation. I conclude by arguing that constitution relations neither exhaust the totality of conditions that a proper whole is subject to, nor do they preclude other types of relations that one proper whole can have in relation to another proper whole.

2. Mechanistic Constitution

Unpacking mechanistic constitution requires understanding 3 main concepts: (2.1) levels of mechanism, (2.2) relevance condition, and (2.3) mutual manipulability. The three concepts are answers to three questions: *what* stands in a constitution relation, *how* do we identify a constitution relation, and *why* is a constitution relation noncausal. Let us address each concept:

(2.1.) Craver (2014) explains that "[i]n levels of mechanisms [...] lower-level components are organized together to make up some behavior or property of the whole" (17). Accordingly, each level is populated by a set of components which, like legos, make up, or *constitute*, a larger component at a level higher than the level they occupy. What makes the components more than just legos is their mere aggregation doesn't result in the functioning of the whole, rather they need to be spatially and temporally connected in specific ways to give rise to the function of the whole. The reason for restricting each level to a set of lower-level components is that it allows the mechanist to identify what a proper whole is in a systematic way. The worry, as Craver puts it, is that without a systematic way of identifying which components belong to which proper whole, we could end up with what Lewis (1991) called "unrestricted composition", which allows for "gerrymandered wholes" to have the same status as proper wholes. By defining the relationship between a proper whole and its components as one of the latter making up, or *constituting*, the first, we have a way of delineating one proper whole from another.

(2.2.) Having established that a proper whole is made up of, or constituted by, its components, a natural question to ask is regarding the features of this relationship of constitution. To put it differently, what features of the lower-level components are we looking for in trying to identify the proper whole which they constitute? Craver (2014) gives

the answer: relevance condition. The idea is that what is relevant to the constitution of a proper whole is the stuff which contribute to its activities. As he puts it, "all the lower-level properties, activities, and organizational features of the parts are relevant to—*contribute to*—the property or activity of the whole" (15, emphasis added). Notice that we are identifying a relation of constitution, the *stuff* which make up one proper whole, by one of contribution to the *activities* of the proper whole. A way to put this clearly is to say that what Craver is after can be represented in the slogan: *what constitutes, contributes*.

(2.3.) Craver insists that a constitution relation should be understood not in causal terms, rather in terms of mutual manipulability: "a component is relevant to the behavior of a mechanism as a whole when one can wiggle the behavior of the whole by wiggling the behavior of the component and one can wiggle the behavior of the component by wiggling the behavior as a whole" (Craver, 2007, 153). His three reasons for advancing this claim are: 1) mechanistic constitution lacks the *distinctness* typical of causal relations. In a mechanistic constitutive picture, there's no way of isolating the effect that wiggling the behavior of the whole would have on the behavior of a single component without it affecting other components. 2) mechanistic constitution is *synchronic*, while "causes must precede their effects". As Craver explains, "if one understands the realization relationship as a synchronic relation, then levels of realization cannot be causally related" (Craver 2014, 11). 3) mechanistic constitution is *bidirectional* while causal relations are unidirectional (Craver 2007, 153).

From this, we get the following picture: mechanistic constitution is understood in terms of *lower-level components*, occupying a level below that where the proper whole they constitute is. The constitution relation is understood in terms of these lower-level components *contribution to the activities* of the proper whole. The constitution relation is *not causal* since it lacks distinctness, is synchronic, and bidirectional, while a causal relation is distinct, asynchronous, and unidirectional. The upshot of this picture is that it allows us to delineate one proper whole from another in a systematic way, without falling into the Lewisian pit of unrestricted composition. This understanding of mechanistic constitution is taken to preclude interlevel causation since the contribution of lower-level components to the activities of the whole is a matter of realization, not causation. I take it that something like this is what Craver and Bechtel meant when they said, "[t]he idea of causation would have

to stretch to the breaking point to accommodate interlevel causes" (547). And if we can reserve the concept of causation to where we have distinct, asynchronous, and unidirectional relations, we can maintain conceptual clarity, avoid incoherent talk of wholes "causing" their parts, and avoid falling prey to strong (spooky) emergence.

While this picture of mechanistic constitution precludes the possibility of constitutive downward causation, Woodward takes a stab at drawing a picture of downward causation compatible with mechanistic constitution. He explains that in considering causal factors, distinctness is a matter of variables, not of entities. Accordingly, while entities cannot stand in causal relations, variables representing wholes and parts, that stand in compositional relations, can themselves stand in causal relations. He gives an example of the Hodgkin–Huxley model where "the channels are "part" of the cell membrane [...] and the behavior of the channels, including their conductances, is influenced by [...] the potential difference *V* across the entire membrane" (2021a, 181). Moreover, Woodward points to the fact that, while manipulating *V*, there is "a very short temporal delay between the momentary value of the membrane potential or its time derivative and the response of the ion channels" (2021b, 232). This means that the relation between the whole and its parts is asynchronous and not instantaneous as Craver conceives of it. Furthermore, Woodward explains that bidirectionality is a matter of causal modeling of feedback loops and equilibrium cycles. He suggests that assigning temporal indices can help distinguish variables from one another, which would in turn allow us to understand when the cycle begins and when it ends. By indexing cyclical causal relations temporally, there's no need to represent it as a cycle, rather it can be represented by specifying the directions, indicating that "X $\rightarrow$ Y is a different relation from Y $\rightarrow$ X" (Woodward 2021b, 234).

Woodward's solution is that rather than speaking of downward causation in terms of wholes standing in a causal relation to their parts, which he agrees is incoherent, we can speak of it in terms of manipulable variables representing factors which stand in causal relation to one another regardless of their constitutive relation. In a way, Woodward's proposal *flattens out* the constitutive picture: whenever we can manipulate variables to get a change in other variables, we have causation. Nevertheless, the hierarchy doesn't disappear if we treat it as flat. One way to think of this is as an epistemic move designed to give us a way to speak of something intuitive, i.e. wholes exerting influence on their parts, without it violating a core

metaphysical assumption, i.e. wholes inability to stand in a causal relation to their parts. Nevertheless, it seems odd to say that we can intervene to manipulate variables representing causal factors, yet these causal factors, which represent parts and wholes, don't stand in a causal relation to one another! For instance, we can intervene to change the potential difference across the membrane, which will lead to changes in the behavior of ion channels, yet we cannot say that the membrane stands in a causal relation to its ion channels only because ion channels are what make up the membrane. Either we have a causal relation between two factors, or we do not. Woodward's proposal forces us to confront an uncomfortable choice: either we accept that what is metaphysical can only be inferred from what is epistemically possible, or we maintain our metaphysical intuitions in the face of empirical evidence. Debating this choice falls outside of the scope of this paper. What I wish to do here is to present further evidence that supports the idea that we can have constitutive downward causation *even when we lack some of the three desiderata* discussed by Craver and Woodward. In the following section, I consider how Ross's (2023) notion of causal constraints provides further challenge to the claim that mechanistic constitution precludes downward causation.

3. Causal Constraints

As Ross explains, constraint-based explanations are often contrasted with causal-mechanical explanations, as two "diverging scientific practices" (Green and Jones 2016, 343). While the first looks at elements which limit a process or a system, the latter looks at elements which trigger or change a process or a system. The traditional way of distinguishing between the two types of explanation is in terms of the first being non-causal while the latter is. The idea is that to constrain a process is not the same as to trigger or change it, and for a relationship to count as a cause, it needs to have an *active dimension* to it. Since constraints are traditionally seen as passive, they do not count among the class of causes under the traditional view, which follows a production or transference view of causation, where for something to count as a cause, it either needs to induce an effect in Y or carry a mark from X to Y.

Ross (2023) challenges this traditional understanding of constraints, arguing that some constraints exhibit features which make them causal under an interventionist account of causation. Ross argues that constraints can count as causal if they involve "a (a) manipulable explanans and a (b) dependency relation that contains empirical information" (15). The idea is that, under an interventionist account, we do not look at causation in terms of productivity or transference, rather in terms of invariance under manipulation, i.e. the effects continue to obtain whenever an intervention occurs. If an intervention occurs on a constraint leading to an effect, and this effect is invariant, then the constraint counts as causal. Examples of causal constraints include riverbanks and metabolic pathways, where an intervention on the first leads to changes to the flow of the river, and an intervention on the latter causes changes to "which types of downstream products the upstream substrate is converted into" (14).

Nevertheless, Ross is aware that constraints are different from triggering causes, even if, under an interventionist account, they both can count as causal. She provides four main features which distinguishes constraining causes from triggering causes:

1- They *limit* the set of possible values an explanatory target can take
2- They are understood to be *external* to the process or system they limit
3- They are *relatively fixed* compared to other factors in the system
4- They *shape* the presentation of an effect, rather than triggering it

Notice the ways in which causal constraints are different from triggering causes: 1) triggering causes take on specific values, while constraints limit the set of possible values a triggering cause may have, 2) triggering causes are internal actors in a process or a system, whereas constraints delimit a process or system by being external to it. 3) triggering causes are variable, whereas constraints are fixed, at least relative to the triggering causes they delimit. 4) triggering causes start a causal chain, whereas constraints shape the trajectory of the ensuing effect. If we apply this schema to Woodward's' membrane/ion channels example, we find that it's a case of a triggering cause: the potential difference *V* across the entire membrane needs to take on a certain value in order for ion channels to behave a certain way, the process is internal to the production of action potentials, the values of the potential difference *V* are variable, and it is necessary to start a causal chain.

More crucially to our purposes, from this picture, it follows that only triggering causes can be distinct (they take on specific values), asynchronous (they precede their effects), and unidirectional (they start a causal chain), while constraining causes lack distinctness (they limit the values of triggering causes), are synchronic (constantly present to shape the effects of triggering causes), and bidirectional (they allow for feedback communication between upstream causes and downstream effects). Recall that, according to the mechanistic picture, a causal relation, similar to triggering causes, is distinct, asynchronous, and unidirectional, while a constitution relation, similar to constraining causes, lacks distinctness, is synchronic, and bidirectional. This symmetry between constraining causes and constitution relations allows us to translate the first into the latter straightforwardly: instead of speaking of lower-level mechanisms *making up* higher-level mechanisms, we can speak of higher-level mechanisms *constraining* lower-level mechanisms. Instead of speaking of lower-level mechanisms *contributing* to the effects of higher-level mechanisms, we can speak of higher-level mechanisms *influencing* the behavior of lower-level mechanisms. From this, it becomes clear that the idea of causation need not stretch to the breaking point to accommodate interlevel causes. *We can have interlevel causation once we recognize that mechanisms at one level are constrained by others at a higher-level just as much as they constitute them*.

As Ross admits, causal constraints have received little attention in the literature. She explains that "[r]easons for this inattention to causal constraints are numerous, complicated, and deserving of serious attention and analysis. The conception of constraints as relatively fixed, unchanging, or difficult to manipulate may lead to a decreased appreciation of their explanatory power" (Ross 2023, 16). However, that they are difficult to manipulate does not mean it is impossible. In the next section, I present an example of a constraining cause, arguing that the relationship it has with the mechanisms it constrains is an instance of constitutive downward causation lacking traditional causal desiderata under an interventionist account of causation.

4. The Case of the Structural Connectome

Weninger et. al (2022) investigated the role of the structural connectome of the brain, as the physical substrate responsible for information propagation, in a) determining "the set of

possible brain states and [b) constraining] the transition between accessible states". They began by identifying an informational state (I_{state}) as "a function of the activation distribution, where statistically rare values of [fMRI] activation correspond to high information content", such as social cognition, and statistically common values of fMRI activation correspond to low information content, such as gambling (1). Then, they identified the cost of transition between different I_{states} as E_{min}, where E_{min} is "the minimum control energy associated with an optimum trajectory between two states" (4). The researchers were able to corroborate that transitioning into a higher I_{state} requires a higher E_{min}, while a lower E_{min} was required to transition into a lower I_{state}.

The researchers then set out to answer whether the architectural features of the structural connectome facilitate transitioning between I_{state}. What they did was that they compared two models of the structural connectome, one representing its architectural features, the other is "a randomized model of brain connectivity that preserved the original degree, weight, and strength distributions" (6). They compared the two models with respect to the mean difference in control energy for 3 I_{states} (low, intermediate, and high). Not only did they find the model representing the architectural features was more cost-effective in transitioning in all three states, but they also found that "it was particularly suited to reach high information content states when compared to a null model" (6).

What we have here is a case where the structural connectome of the brain acts as a constraining cause to the causal relation between I_{state} and E_{min}. The structural connectome:

1) sets the possible values of the cost to transition between I_{states}
2) is external to I_{state}
3) is relatively fixed compared to process of transitioning from one I_{state} to another
4) shapes, rather than triggers, I_{states} transitions

The reason it is a constraining cause is that the efficiency of state transition counterfactually depends on certain architectural features it exhibits. Thus, regardless of the mechanisms responsible for giving rise to I_{states} and E_{mins}, it is the architecture of the structural connectome that provides an explanation as to why the brain is suited for reaching high information content states. More crucially, changes in these architectural features entail

changes in the efficiency of transitioning between different I_{states}. Given that the processes of the brain are sensitive to changes in the structural connectome, and the structural connectome exists at a higher level from the mechanisms which give rise to I_{states} and E_{mins}, it follows that what we have here is a case of constitutive downward causation, whereby a higher-level entity exerts some influence on lower-level mechanisms by virtue of being a causal constraint on them.

Aside from any metaphysical considerations, such a research program is both exciting and sufficiently explanatory. Weninger et. al's work corroborated the findings from Sorrentino et. al, who found "the structural connectome relates to, and likely affects, the rapid spreading of neuronal avalanches" (2021, 1). These findings prompted Schiff (2022) to ask "could we use such an information-control theoretic description of brain activity and mental effort to shed new light on cognitive dysfunction and mental health? An information-based dynamic biomarker for cognitive disorders would be very useful if it were to emerge from such a framework" (2022). The potential therapeutic benefits of such a research program far outweighs the debate around whether we should characterize the relationship between the structural connectome and brain dynamics as causal or mechanistically mediated.

5. Conclusion

The purpose of this paper is to advance an argument from causal constrains that one can talk of constitutive downward causation without it being incoherent, sounding awkward or unnatural, or entailing strong emergence. The reason why it makes no sense on a mechanistic picture is that the condition of constitution requires that contribution to the activities of a mechanism *only* comes from the mechanisms which make it up. Such a reductive picture neglects the various contributions which could come from the sides or above to regulate and influence the behavior of a mechanism at any given level. Earlier, I said that what Craver is after can be represented in the slogan: *what constitutes, contributes*. However, mechanists' true slogan is that *'only' what constitutes, contributes*; and this is simply false. Sound scientific practices point us towards the utility of such interventionist approaches at higher-level constrains in order to track effects on lower-level mechanisms. While this paper cannot claim to have settled the debate on whether downward causation

is a matter of variables or entities, what it did contribute is a novel argument showing that we still can make sense of causal claims even if we lack traditional causal desiderata such as distinctness, asynchronousity, and unidirectionality.

What this also shows is that philosophers need to be careful about making grand claims regarding the nature of reality. While neat, the idea of levels of mechanisms only serves to illustrate the utility of a certain relevance condition (i.e. constitution) in performing a certain task (i.e. demarcate proper wholes). It does not follow from it that this relevance condition, which only serves one task, either *exhausts the totality* of conditions that the proper whole is subject to or *precludes* any other type of relation that a proper whole can have in relation to another proper whole. Such determination of relations is subject to scientific investigations which are ongoing and evolving. Yesterday's mechanistically mediated effects may turn out to be tomorrow's causal constraints. If the absence of causal relations between levels is needed to demarcate one level from another, then this should put the very notion of levels in question before we start questioning causal relations among levels.

References

Craver, Carl. 2014. "Levels." In *Open Mind*. Open MIND. Frankfurt am Main: MIND Group.

Craver, Carl. 2007. *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Clarendon Press.

Craver, Carl, and William Bechtel. 2007. "Top-down causation without top-down causes." *Biology & philosophy* 22: 547-563.

Eronen, Markus. 2021. "Levels, Nests, and Branches: Compositional Organization and Downward Causation in Biology". In *Levels of organization in the biological sciences*, edited by Brooks, D. S., DiFrisco, J., & Wimsatt, W. C. 77-87. Cambridge, MA: The MIT Press.

Green, Sara, and Nicholaos Jones. 2016. "Constraint-based reasoning for search and explanation: Strategies for understanding variation and patterns in biology." *dialectica* 70, no. 3: 343-374.

Ross, Lauren. 2023. "The explanatory nature of constraints: Law-based, mathematical, and causal." *Synthese* 202, no. 2: 56.

Schiff, Steven. 2022. "In the Brain, Function Follows Form." Physics, July 5. https://physics.aps.org/articles/v15/93.

Sorrentino, Pierpaolo, Caio Seguin, Rosaria Rucco, Marianna Liparoti, Emahnuel Troisi Lopez, Simona Bonavita, Mario Quarantelli, Giuseppe Sorrentino, Viktor Jirsa, and Andrew Zalesky. 2021. "The structural connectome constrains fast brain dynamics." *Elife* 10: e67400.

Weninger, Leon, Pragya Srivastava, Dale Zhou, Jason Z. Kim, Eli J. Cornblath, Maxwell A. Bertolero, Ute Habel, Dorit Merhof, and Dani S. Bassett. 2022. "Information content of brain states is explained by structural constraints on state energetics." *Physical Review E* 106, no. 1: 014401.

Woodward, James. 2021a. "Downward causation and levels". In *Levels of organization in the biological sciences*, edited by Brooks, D. S., DiFrisco, J., & Wimsatt, W. C. 175-194. Cambridge, MA: The MIT Press.

Woodward, James. 2021b. "Downward Causation Defended." In *Top-Down Causation and Emergence*, edited by J Voosholz and M Gabriel, 217-251. Cham: Springer International Publishing.