

Making Sense of Interlevel Causation in Mechanisms from a Metaphysical Perspective

Beate Krickel

ABSTRACT

According to the new mechanistic approach, an acting entity is at a lower mechanistic level than another acting entity if and only if the former is a component in the mechanism for the latter. Craver and Bechtel (2007) argue that a consequence of this view is that there cannot be causal interactions between acting entities at different mechanistic levels. Their main reason seems to be what I will call the *Metaphysical Argument*: things at different levels of a mechanism are related as part and whole; wholes and their parts cannot be related as cause and effect; hence, interlevel causation in mechanisms is impossible. I will analyze this argument in more detail and show under which conditions it is valid. This analysis will reveal that interlevel causation in mechanisms is indeed possible, if we take seriously the idea that the relata of the mechanistic level relation are acting entities and accept a slightly modified notion of a mechanistic level that is highly plausible in the light of the first clarification.

1. Introduction

The new mechanistic approach is not only fruitful with regard to methodological and epistemic issues in the philosophy of science (like, for example, scientific explanation and discovery). It also provides a good starting point for a fruitful metaphysics of the special sciences—a metaphysics that starts with a descriptively adequate analysis of the epistemology of the sciences, investigates the ontological commitments¹ of it, and aims at building the fundament for a conceptually and metaphysically kosher philosophy of the special sciences. Some authors explicitly deal with metaphysical questions that are crucial to the mechanistic approach of scientific explanation. For example, Carl Craver defends the so-called *ontic view* of explanation according to which mechanisms exist mind-independently and are the truth-makers of mechanistic explanations (Craver 2007; Craver 2014); Phyllis Illari and Jon Williamson (2013) argue that a mechanistic ontology has to include entities and activities, where the latter are supposed to account for the causal aspects of mechanisms; Stuart Glennan (2010) argues that the metaphysics of causation should be understood in terms of mechanisms; and Carl Gillett

¹ Following Quine (1948), one might distinguish between metaphysics and ontology by assuming that ontology is about what there is (Which kinds of things do we have to assume?), whereas metaphysics addresses the features, relations etc. between these things (e.g. What is causation if we take it to be a relation between these things?).

(2013; 2010) argues against traditional accounts of realization on the basis of his analysis of the making-up relations in mechanisms.

Analyzing the metaphysical implications of the new mechanistic approach is particularly interesting when they touch classical metaphysical problems like those discussed in the philosophy of mind literature. One metaphysical problem central to the philosophy of mind and the metaphysics of science in general is the question of whether there can be causal relations between entities at different levels. For example, one might hold that mental causation is a case of downward causation (Robb and Heil 2014). If one could make sense of downward causation in the mechanistic framework, this might provide a fruitful basis for thinking about mental causation. Answering traditional metaphysical philosophical questions like those concerning mental causation based on a metaphysical analysis of the new mechanistic approach is especially promising because it guarantees an answer that is rooted in an empirically justified approach.

The purpose of the present paper is to analyze the metaphysics of (apparent) interlevel causation in mechanisms. Contrary to Craver and Bechtel (2007), I will show that the new mechanistic approach is indeed compatible with a literal understanding of interlevel causal claims. After clarifying the metaphysical commitments of the new mechanistic approach with regard to the notion of levels of mechanisms, it will become clear that there is room for downward and upward causation between mechanistic levels. The paper proceeds as follows: in Section 2, I will present the notion of a mechanism and that of a mechanistic level as introduced by Craver (2007). In Section 3, examples of apparent interlevel causation are introduced and the main objection against a literal understating of interlevel causal claims, which I will call the *Metaphysical Argument*, will be presented. In Section 4, I will provide a more detailed analysis of the *Metaphysical Argument*. I will argue that it is unclear what the relata of the part-whole relation are on which the *Metaphysical Argument* hinges. According to my analysis, the relata are best characterized as what I will call *entity-involving occurrents* (EIOs). In Section 5, I will show that on the basis of these insights, it is not only possible to illuminate the part-whole relation involved in mechanisms. We can also show that interlevel causation in mechanisms is possible if one accepts a slight but highly plausible modification of the definition of a mechanistic level.

2. Levels of Mechanisms

The new mechanists argue that mechanisms come in hierarchies of *levels of mechanisms*. This is clearly a metaphysical notion, since according to Craver, levels of mechanisms are “features of the world rather than (...) features of the units or products of science“ (Craver 2007, 177). The relata at the different levels are taken to be behaving mechanisms (or “phenomena”) and their component entities and activities (Craver 2007, 189). Craver illustrates the notion of a mechanism, and that of a mechanistic level, with the help of what has come to be known as the *Craver-diagram* (see Figure 1).

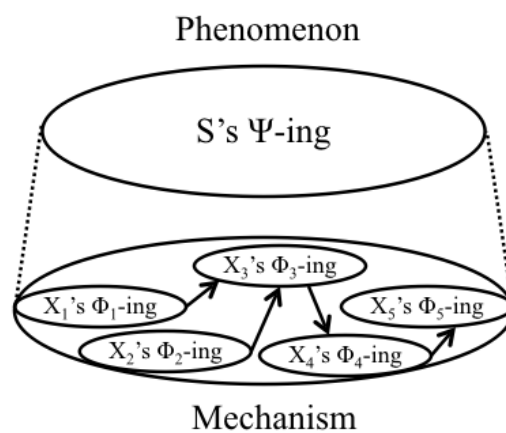


Figure 1 Mechanisms according to Craver (adapted from Craver (2007, 7))

In Figure 1, the behaving mechanism, which is also referred to as ‘phenomenon’ is at the higher level.² The behaving mechanism/the phenomenon is represented by ‘S’s ψ -ing,’ where ‘S’ refers to the mechanism as a whole, and ‘ ψ ’ represents the property or behavior of S that is to be explained. According to Craver, the phenomenon is explained by the organization of its component entities and activities that make up the mechanism at the lower level (Craver 2007, 7). In Figure 1, the Xs represent the entities that are components of the mechanism, and the ϕ s are the activities performed by these entities. The arrows between the Xs’ ϕ_n -ings indicate that the components are causally interacting. The dotted lines indicate the hierarchy between the levels. But how do we have to understand this level-talk exactly? Let us first take a look at an example. One prominent example are the different levels of spatial memory (Craver 2007, 165–70). Figure 2 provides an overview of the different mechanistic levels of spatial memory.

² Note that there is some ambiguity in the mechanistic literature, whether the phenomenon just is the *behaving mechanism*, whether it is a *behavior* of the mechanism, or whether ‘S’ refers to a system that *contains* the mechanism (see Krickel (under review)). For the sake of argument, I will ignore this ambiguity here, and assume that the phenomenon is a behaving mechanism.

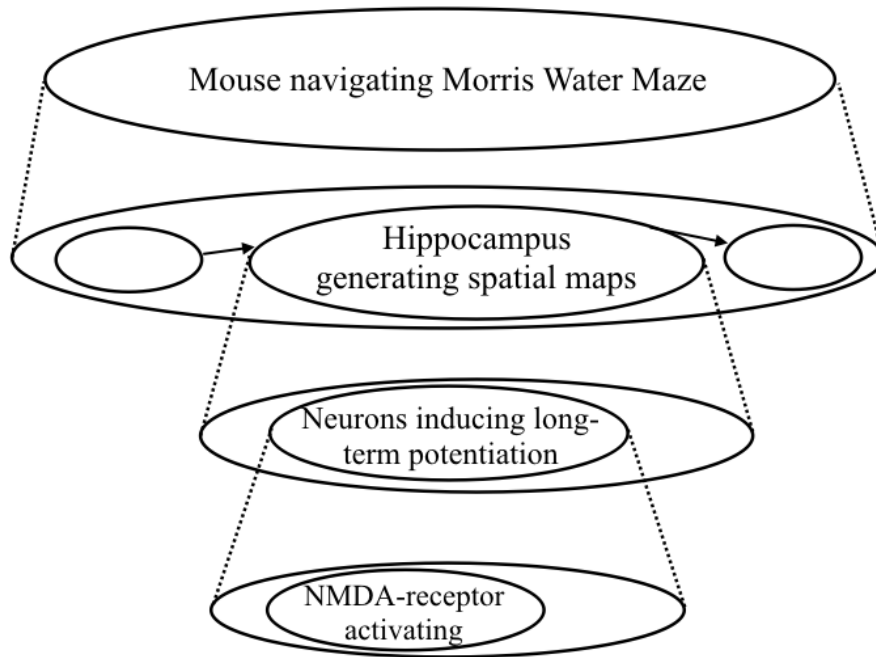


Figure 2 Example of different level of mechanisms (loosely adapted from (Craver 2007, 166))

As Figure 2 shows, at the top level of spatial memory is a mouse navigating the Morris Water Maze, which is supposed to be an instance of spatial memory behavior. One component of the mechanism that is responsible for the mouse's navigation behavior is the hippocampus that generates spatial maps, which is therefore at a lower level than the Mouse navigating the Morris Water Maze. The hippocampus can generate spatial maps because it consists of neurons that induce long-term potentiation. Hence, the neurons producing long-term potentiation are at a lower level than the hippocampus generating spatial maps. Again, the long-term potentiation is enabled by NMDA-receptors, which are, thus, at a lower level than the neurons inducing long-term potentiation. More formally, Craver defines levels of mechanisms as follows:

(Levels of Mechanisms) X's ϕ -ing is at a lower mechanistic level than S's ψ -ing if and only if X's ϕ -ing is a component of the mechanism for S's ψ -ing. (Craver 2007, 189)

The relata X's ϕ -ing and S's ψ -ing are thereby taken to be "acting entities" (Craver 2007, 189). An acting entity X's ϕ -ing is a component of a mechanism for S's ψ -ing, according to Craver's mutual manipulability account, if the following holds:

(Mechanistic Component) X's ϕ -ing is a component in the mechanism for S's ψ -ing if and only if:

- i. X is part of S, and
- ii. X's ϕ -ing and S's ψ -ing are mutually manipulable. (Craver 2007, 153)

Condition i. specifies that X has to be a spatiotemporal part of S: the entity-component (X) has to be a spatiotemporal part of the system whose behavior is to be explained. Condition (ii) is spelled out in terms of interventionism (Woodward, 2003, 2015): X's ϕ -ing and S's ψ -ing are mutually manipulable if and only if it is possible to ideally intervene into X's ϕ -ing and thereby change S's ψ -ing, and it is possible to ideally intervene into S's ψ -ing and thereby change X's ϕ -ing. An intuitive understanding of interventions suffices for present purposes.³ Bottom-up manipulability of the phenomenon by intervening into a component of the mechanism alone is not sufficient to establish constitutive relevance. Craver introduces *mutual* manipulability in order to be able to distinguish between parts that are components and those that are mere background conditions, such as the heart's beating which is a background condition of almost all mechanisms in living organisms. While it is possible to, for example, change word stem completion behavior by intervening into the heart's beating, according to Craver, it is not possible to top-down manipulate the heart's beating by intervening into word-stem completion behavior (Craver 2007, 158).⁴

One can illustrate this formal characterization with help of the above example (Figure 2): The hippocampus generating spatial maps is at a lower level than the mouse navigating the Morris Water Maze because the former is a component of the mechanism for the latter. It is a component because the hippocampus is a part of the mouse (condition i. of *Mechanistic Component*), and the hippocampus's activity and the mouse's activity are mutually manipulable (condition ii. of *Mechanistic Component*; by changing the hippocampus's activity you can change the mouse's behavior, and vice versa). The same holds for the other lower mechanistic levels: each acting entity is a

³ Many authors argue that the mutual manipulability account and interventionism are incompatible (Leuridan 2012; Baumgartner and Gebharder 2015; Casini and Baumgartner 2016; Romero 2015). Therefore, it remains controversial how to understand the claim that phenomena and mechanisms mutually depend on each other (note that promising attempts have been made to save the combination between constitutive explanation and interventionism; see (Baumgartner and Gebharder 2015; Casini and Baumgartner 2016; Krickel (forthcoming)). Here, I will ignore these problems for the sake of argument. Furthermore, as I will argue below, since interventionism is not a metaphysical account, it might be possible to accept the incompatibility of interventionism and mutual manipulability, while keeping the mutual manipulability criterion by providing a metaphysical analysis of manipulability (e.g., in counterfactual terms along Lewisian lines).

⁴ However, Craver admits that further pragmatic considerations might be necessary to draw the line between background conditions and components (Craver 2007, 157).

part of the entity at the higher level, and the acting entities at the different levels are mutually manipulable.

Based on this notion of a mechanistic level, Craver and Bechtel (2007) have argued that causal interactions *between* levels are impossible. I will discuss their main argument for this deny of interlevel causation in mechanisms in the following section.

3. Interlevel Causation & The Metaphysical Argument

Craver and Bechtel (2007) discuss the case of a general who is killed by a heart attack as an example of (putative) *bottom-up causation* between mechanistic levels:

If the general is viewed as a living human being (as opposed to an occupant of an office or rank in the military) whose behaviors are those of ordinary life (locomoting, communicating, etc.), then the heart is a component in one of her bodily mechanisms (e.g., the circulatory system). The failure of blood supply to the heart muscle is a failure of a part of the overall mechanism, and the ‘effect’ is on the whole mechanism – namely, death. (Craver and Bechtel 2007, 557)

The heart attack (the putative cause) is a component in at least some mechanisms responsible for the general’s living-behaviors (locomoting, communicating, etc.). Hence, the heart attack is at a lower level than the general’s living behaviors. Since the effect of the heart attack (death) concerns the disruption of these living-behaviors, this seems to be a case of bottom-up causation. An example of putative *top-down causation* in a mechanism is the following:

Hal steps onto the court, serves, and so begins the tennis match. Very quickly, blood borne glucose is taken up through the cell membrane. Once inside, it is phosphorylated and bound into molecules of hexosediphosphate (...) Hal’s tennis playing also alters the behavior of innumerable biochemical pathways and cellular mechanisms that are involved in his tennis playing, both in the short-term and in the long-term. Why did Hal’s cells start using more glucose (i.e., binding glucose into molecules of hexosediphosphate)? Because Hal started to play tennis. Similar stories could be told about Hal’s respiratory mechanisms, visual system, and many others besides. Changing the behavior of the mechanism as a whole changed the activities of its components. (...) This is the sort of case for which appeal to top-down causation seems most compelling. (Craver and Bechtel 2007, 559)

Here the apparent interlevel causal relation holds between a person playing tennis and her cells consuming glucose. The behavior of the cells is at a lower mechanistic level than the person playing tennis because the former is a component of the mechanism that is responsible for the person’s tennis playing.

It has been argued by different authors that the notion of a mechanistic level does not allow for interlevel causation (Craver 2007; Craver and Bechtel 2007; Bechtel 2008;

Baumgartner and Gebharder 2015; Casini and Baumgartner 2016; Romero 2015; Gebharder 2015). Hence, it is argued, the putative cases of bottom-up and top-down causation indeed cannot be given a strict causal reading. The heart attack is not a cause of the general's death and the tennis playing is not a cause of the cells' glucose consumption. One reason for this denial of interlevel causation in mechanisms stems from the fact that the interlevel relation (mutual manipulability) is spelled out in terms of interventionism. Michael Baumgartner and Alexander Gebharder (2015), and Felipe Romero (2015) argue that interventions into higher-level phenomena are necessarily *fat-handed* in the sense that they necessarily change the higher-level phenomenon and at the same time one of the components of the mechanism at the lower level. Hence, these interventions are not ideal and thus cannot establish a causal relation between the two levels. If one takes the existence of ideal interventions to be a necessary condition for causation, it follows that there cannot be causal relations between levels.

From a metaphysical perspective, one might doubt whether the non-existence of ideal interventions into higher-level phenomena proves that there cannot be causal relations between levels. Interventionism is not a metaphysical approach to causation and, thus, it might be that in fact there is interlevel causation but we are unable to establish it by means of interventions. Still, authors such as Craver and Bechtel (2007) have metaphysical reasons to deny interlevel causation in mechanisms as well.⁵ The argument seems to be, roughly, that the concept of causation implies that causes and effects are wholly distinct, whereas the entities at different levels of mechanisms are related as part and whole (Bechtel 2008, 153–55; Craver 2007, 153–54, 195; Craver and Bechtel 2007, 551–55). Hence, they cannot be causes or effects of one another. More concretely, the following *Metaphysical Argument* seems to be what underlies Craver's and Bechtel's line of argument:

(The Metaphysical Argument)

⁵ A further metaphysical reason to deny interlevel causation in mechanisms that is independent of the interventionist approach is the causal exclusion argument (Gebharder 2015; Romero 2015; Kim 2005). I will not address this objection here because, first, the objection discussed in this paper is more fundamental than objections based on exclusion worries. The present objection, roughly, is that the level-relata do not even qualify as possible candidates for causal interactions. The exclusion argument starts with the assumption that things at different levels, *prima facie*, are candidates for being causes and effects. Hence, if it turned out that the objection discussed in this paper goes through, there is no need to talk about exclusion worries at all because mechanistic interlevel causation turns out to be impossible for more basic reasons. Of course, if it turns out that the objections discussed here can be rejected, one has to meet the challenge posed by the exclusion argument. To see how this could be done in the context of mechanistic levels see Krickel (under review).

1. X's ϕ -ing is at a lower mechanistic level than S's ψ -ing iff X's ϕ -ing is a component of the mechanism for S's ψ -ing. [def.]
2. Interlevel causation is a causal relation between things at different levels. [def.]
3. Mechanistic interlevel causation is possible iff there can be causal interactions between a component of the mechanism for S's ψ -ing and S's ψ -ing. [from 1, 2]
4. X's ϕ -ing is a component in the mechanism for S's ψ -ing iff X is part of S, and X's ϕ -ing and S's ψ -ing are mutually manipulable. [def.]
5. If X's ϕ -ing is a component in the mechanism for S's ψ -ing, then X's ϕ -ing and S's ψ -ing are related as part and whole. [from 4]
6. Mechanistic interlevel causation is possible only if there can be causal interactions between a whole and its parts. [from 3, 5]
7. Cause and effect are wholly distinct, their occurrence is temporally asymmetric, and effects depend⁶ on causes, but not *vice versa*.⁷ [def.]
8. Wholes and their parts are spatiotemporally coexisting or at least overlapping, changes in them occur simultaneously, and they are mutually dependent. [def.]
9. Things that are related as part and whole cannot be related as cause and effect, and *vice versa*. [from 7, 8]
10. Mechanistic interlevel causation is not possible. [from 6, 9]

Premise 1 is a statement of Craver's definition of levels of mechanisms (see Section 2). Premise 2 is a rather trivial fact about interlevel causation. Statement 3 follows from 1 and 2: mechanistic interlevel causation is possible if and only if there can be causal interactions between a component of a mechanism and the phenomenon that is brought about by that mechanism. According to the definition of what it is to be a mechanistic

⁶ It is not entirely clear how Craver and Bechtel understand the difference in dependency between causes and effect, on the one hand, and wholes and their parts, on the other. This is what they write: „It is a widely accepted condition on accounts of causation that they account for the asymmetry of causal dependency. The sun's elevation causes the length of the shadow, but the length of the shadow does not determine the elevation of the sun. The virus produces the spots on the skin, but the spots on the skin do not cause the infection with the virus. [...] While at least some cases of intralevel causation are asymmetrical, all of the interesting cases of interlevel causation are symmetrical: components act as they do because of factors acting on mechanisms, and mechanisms act as they do because of the activities of their lower-level components.“ (Craver and Bechtel 2007, 553)

⁷ One might contest that causation and the part-whole relation indeed have the features assumed in premises 7 and 8. For the sake of argument I will accept the truth of these premises. I take these statements to be summaries of Craver's and Bechtel's considerations.

component introduced in Section 2 (premise 4), the components of a mechanism and the phenomenon that is brought about by that mechanism are related as part and whole (statement 5). Hence, in order for there to be mechanistic interlevel causation, it must be possible that wholes and their parts can causally interact (statement 6). But, as stated in 9, it is commonly accepted that causes and effects, on the one hand, and wholes and their parts, on the other, have characteristics that renders causal interactions between wholes and their parts impossible (premises 7 and 8). Hence, it follows that interlevel causation between things at different levels of mechanisms is impossible (10).

Denying the truth of claims such as ‘The heart attack is the cause of the general’s death’ or ‘Hal’s playing tennis causes his cells to consume more glucose’ seems to be odd. Claims like these do not only reflect our everyday talk about causal relations but also scientific causal claims. Craver and Bechtel admit this:

Many philosophers (e.g., Alexander 1927; and several authors in Andersen et al. 2000) and scientists (e.g., Morgan 1927; Campbell 1974; Sperry 1976) appeal to top-down causes in their explanations. (Craver and Bechtel 2007, 547)

In order to, on the one hand, account for the fact that philosophers and scientists are not speaking falsely every time they make interlevel causal claims, while on the other hand, account for the impossibility of interlevel causation in mechanisms, Craver and Bechtel develop an account of what they call *mechanistically mediated effects* (Craver and Bechtel 2007). According to this account, causation occurs only intralevel. The interlevel relation is taken to be a non-causal constitutive relation. In order to see how this account reduces interlevel causation to mechanistically mediated effects, consider the case of apparent top-down causation. According to Craver and Bechtel, top-down causation indeed consists of a constitution relation between the apparent higher-level cause and a lower-level mechanism, and a causal relation between the lower-level mechanism and the lower-level effect. The effect of the apparent higher-level cause, indeed, is an effect of the lower-level mechanism that constitutes it. Hence, according to this analysis, literally speaking, interlevel causal claims are false.

I will show that the mechanistic approach indeed provides resources to account for the *literal* truth of interlevel causal claims such as ‘The heart attack is the cause of the general’s death’ or ‘Hal’s playing tennis causes his cells to consume more glucose’ without “evok[ing] concerns that the notion of top-down causation is incoherent or that it involves spooky forces exerted by wholes upon their components” (Craver and Bechtel 2007, 548). A literal interpretation of interlevel causal claims is more

straightforward than the interpretation based on mechanistically mediated effects since it has to postulate only one relation (causation between behaving mechanisms/phenomena and their components) instead of two (causation between components of mechanisms, and constitution between mechanistic components and behaving mechanisms/phenomena) in order to account for the truth of interlevel causal claims. Furthermore, as I will show, the metaphysical tools one needs to make sense of interlevel claims in a literal way is already inherent to the new mechanistic approach. No additional metaphysical claims are necessary that the new mechanists do not already accept, or are committed to—all one has to do is to lay bare the metaphysical commitments of the new mechanists' claims regarding levels of mechanisms and their relata (acting entities).

In the following section I will first show that the Metaphysical Argument as it is presented above is not valid. However, there is an easy fix, which becomes obvious if one takes a closer look at the metaphysics of the relata of the mechanistic interlevel relation. Still, the analysis of the Metaphysical Argument will show that there is a way to make sense of interlevel causation in mechanisms.

4. Clarifying the Metaphysical Argument: the Part-Whole Relation and its Relata

Statement 5 of the Metaphysical Argument states that if X 's ϕ -ing is a component of the mechanism for S 's ψ -ing, then X 's ϕ -ing and S 's ψ -ing are related as part and whole. This is supposed to follow directly from Craver's definition of mechanistic componency (premise 4). But does it follow? How do we have to understand the first condition of Craver's definition of mechanistic componency ("X is part of S")? Leuridan (Leuridan 2012) suggests the following definition of the part-whole relation that Craver might have in mind when defining the first condition for mechanistic componency:

(STI) X is part of S iff the spatio-temporal region occupied by X is contained in the spatio-temporal region occupied by S . (Leuridan 2012)

Does (STI) validate the inference from Craver's definition of mechanistic componency to statement 5? After all, (STI) speaks of part-whole relations that obtain between X and S , which stand for *entities* (material objects). Given this, it is not immediately clear whether the definition also applies to things such as X 's ϕ -ing and S 's ψ -ing (i.e. acting

entities) that statement 5 talks about. Hence, if we interpret the first condition of Craver's definition of mechanistic competency in terms of (STI) it is at best unclear whether statement 5 follows or not because it does not talk about acting entities.

One might argue that the part-whole relation between the entities X and S is sufficient to prohibit causal relations between X and S, and this is sufficient to show that mechanistic interlevel causation is impossible. One problem with this idea is that causal interactions between X and S are precluded for trivial reasons: X and S are simply not the kind of things that can be related by causation. In the literature on the metaphysics of causation, it is a common assumption that causation is a relation between events—not entities (objects). Thus, the relevant question is whether there can be causal interactions between an event in which X is involved (an “X-event”) and one in which S is involved (an “S-event”) given that X is a part of S in the sense of (STI). According to Lewis (D. K. Lewis 1973; D. K. Lewis 1986; D. K. Lewis 2000), events must be wholly distinct in order to be causally related (see premise 7 of the Metaphysical Argument). “Wholly distinct,” roughly, means that the two events occupy different space-time regions. Can X-events and S-events be wholly distinct given that X is a spatial material part of S? Yes, they can: Although X and S are related as part and whole, X- and S-events might occur in different space-time regions. Consider, for example, my going for a walk this morning and my stomach's digesting this afternoon. My stomach is a spatial part of me. Still, my walk and my stomach's digesting are wholly distinct events: My walk this morning occupied a particular route through the park from 9 to 10 a.m.; my stomach's digesting occupied a space-time region inside me while I was sitting on my couch from 2 to 3 p.m. Hence, my walk and my stomach's digesting occur at different space-time regions. Thus, (STI) alone cannot justify the rejection of mechanistic interlevel causation.⁸

Is there another way to spell out the part-whole relation? As already mentioned above, Craver points out that the relata of the mechanistic level relation are not mere material objects/entities. Nor are the relata activities (Craver 2007, 189). Rather, Craver argues, the mechanistic level relation is a relation between “acting entities” (ibid.). Plausibly, Craver takes X's ϕ -ing and S's ψ -ing to stand for acting entities. So maybe the part-whole relation between acting entities (in contrast to the part-whole relation between entities only) prohibits interlevel causation in mechanisms. But what are acting entities and in which sense can they be related as part and whole?

⁸ Note, that when introducing (STI) it is not Leuridan's aim to reject interlevel causation in mechanisms. Rather, he seems to embrace this idea.

Most plausibly acting entities are *events*. Unfortunately there are different conceptions of what exactly events are (for an overview see (Casati and Varzi 2015)). In order to avoid confusions, I will not use the term ‘event’ as an analysis of what acting entities are. Rather, I will use the term *entity-involving occurrent* (EIO) (see also BLINDED). This notion is supposed to make explicit that for the individuation of a particular EIO an entity is crucial (the term ‘event’ is ambiguous in this respect). Furthermore, talking about occurrents in the present context accounts for the idea that “activities exist only extended in time” (Illari and Williamson 2013, 72), while avoiding the problems afflicting the notion of an activity. First, the distinction between occurrents and continuants is common in metaphysics (Simons and Melia 2000). While continuants (such as entities/objects) are said to exist *in* time, occurrents (such as processes) are taken to exist *through* time. Second, by introducing the notion of an occurrent one is not committed to the highly controversial and unclear assumption that the components of mechanisms are *active* or *productive* (Psillos 2004). Occurrents need not be active or productive. Furthermore, the notion of an EIO is supposed to make explicit that, on the one hand, in mechanisms entities are necessarily engaged in a certain behavior, and on the other hand, behaviors cannot occur without an entity that behaves. This is what inspired so-called “entity activity dualism” which is defended by many new mechanists (Machamer, Darden, and Craver 2000). Talking merely about “entities *and* activities” does not express this kind of dependency between entities and activities. Finally, describing the relata of the part-whole relation as EIOs has the consequence that causal relations between wholes and parts are not precluded for trivial reasons. EIOs are plausible candidates for being causal relata (if one agrees that events can be causes and effects, one should be willing to accept that EIOs are, too).

Now, let us come back to the question of which part-whole relations we need in order to get the Metaphysical Argument to go through. EIOs can be regarded as space-time worms. With this picture in mind, we can see that EIOs have parts in two different senses: first, EIOs have, what I will call, *spatial EIO-parts* (cut the space-time worm parallel to the time axis; see Figure 3 a)). Second, EIOs have *temporal EIO-parts* (cut the space-time worm orthogonal to the time axis; see Figure 3 b)). Spatial as well as temporal EIO-parts of EIOs are again EIOs (therefore I speak of “EIO-parts”).

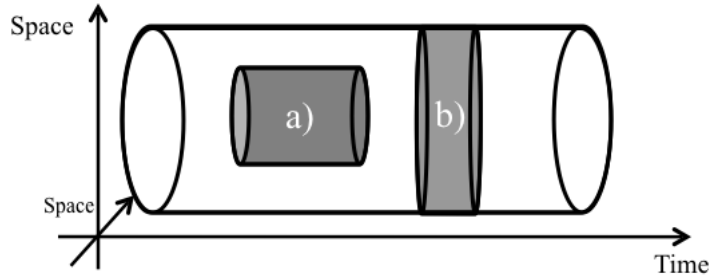


Figure 3 Illustration of a) a spatial EIO-part of an EIO, and b) a temporal EIO-part of an EIO.

Somewhat more formally, the two kinds of part-whole relations between EIOs can be defined as follows:

(spatial EIO-part) EIO_1 is a *spatial EIO-part* of EIO_2 iff:

- (i) the entity involved in EIO_1 occupies a proper spatiotemporal sub-region of the region occupied by the entity involved in EIO_2 ,
- (ii) the occurrent involved in EIO_1 takes place during the occurrence of EIO_2 .

Note that condition (i) of *(spatial EIO-part)* just is the right-hand side of (STI) (see above). Temporal EIO-parts of an EIO are defined as follows:

(temporal EIO-part) EIO_1 is a *temporal EIO-part* of EIO_2 iff:

- (i) the entity involved in EIO_1 is identical with the entity involved in EIO_2 ,
- (ii) EIO_1 begins later and ends earlier than EIO_2 , or EIO_1 begins simultaneously with EIO_2 and ends earlier than EIO_2 , or EIO_1 begins later than EIO_2 and ends simultaneously with EIO_2 .

Consider the following example: Suppose I go for a walk and, at the same time, my stomach is digesting. In this case, the stomach's digesting is a *spatial* EIO-part of me going for a walk. In contrast to that, my turning right to enter the park is a *temporal* EIO-part of my walk. It is important to note the following: the notion of a temporal EIO-part does not presuppose or imply some kind of 4-dimensionalism with regard to objects. I do not argue that my walk through the park involves only a temporal part of me. I am still a three-dimensional object that is wholly present at any time point during my walk. Indeed, I am not concerned with temporal parts of objects here at all. Temporal EIO-parts *are parts of EIOs*, not of objects. In fact, temporal EIO-parts *consist of three-dimensional objects* (that behave in a certain way).

Now, let us apply these insights about EIOs to the notion of a mechanism. The phenomenon S's ψ -ing is an EIO. The mechanism itself is a collection of organized

EIOs (the X s' ϕ -ings) that are *spatial* EIO-parts of the phenomenon-EIO (where being a spatial EIO-part is necessary but not sufficient for being a component of a mechanism; the EIOs and the phenomenon also have to be mutually manipulable (see Section 2)). On the basis of the EIO-interpretation of the relata of the part-whole relation, a mechanism can be illustrated as depicted in Figure 4.

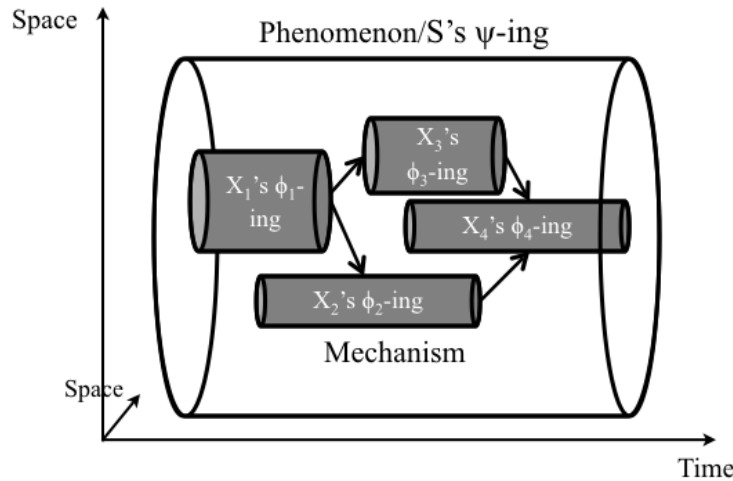


Figure 4 Mechanisms according to the view that the relata of the part-whole relation are EIOs

Here, similar to Craver's illustration (see Figure 1) 'S's ψ -ing' (labeling the big tube) stands for the phenomenon, for example, the dying general, or a person playing tennis. The different X s' ϕ -ings represent the components (the smaller grey tubes), for example, the heart's malfunctioning, or cells consuming glucose. As before, the arrows between the components symbolize causal interactions between them. The organized collection of the components makes up the mechanism. The major difference between Figure 4 and Craver's illustration is that Figure 4 clarifies the temporal and spatial dimensions within a mechanism. The left-right direction is interpreted as the temporal direction, whereas the other two dimensions are spatial dimensions.

On the basis of the EIO-interpretation we can show that the Metaphysical Argument is valid. The insight that components of mechanisms are spatial EIO-parts of phenomenon-EIOs renders the step from premise 4 to statement 5 valid. Premise 4 has to be modified as follows:

- 4.* X 's ϕ -ing is a component of the mechanism for S's ψ -ing iff X 's ϕ -ing is a *spatial EIO-part of S's ψ -ing*, and X 's ϕ -ing and S's ψ -ing are mutually manipulable. [def.]

If we replace premise 4 with premise 4*, statement 5 follows: if X 's ϕ -ing is a component in the mechanism for S's ψ -ing, then X 's ϕ -ing and S's ψ -ing are related as

part and whole. Illuminating the nature of X's ϕ -ing and S's ψ -ing in terms of the EIO-interpretation makes clear why there cannot be causal relations between X's ϕ -ing and S's ψ -ing. The reason is that, if X's ϕ -ing and S's ψ -ing are EIOs and X's ϕ -ing is a spatial EIO-part of S's ψ -ing, then X's ϕ -ing and S's ψ -ing occur in the same space-time region. The notion of causation requires causes and effects to occur in different space-time regions. Hence, there cannot be causal relations between X's ϕ -ing and S's ψ -ing. Note that the same applies to *temporal* EIO-parts: if an EIO is a temporal EIO-part of another EIO, there cannot be a causal relation between them. Again, the reason is that one EIO occupies a spatiotemporal sub-region of the region where the other EIO occurs.

I am not claiming that this is not what Craver and Bechtel actually have in mind. Rather, my claim is that this is in line with and illuminates their ideas. It is plausible that Craver and Bechtel take acting entities to be extended in space *and* time in the sense I explicated above, and that they take the part-whole relation to be a relation between *acting* entities, rather than just entities.

Based on the EIO-interpretation we can show much more. First, one can show that Craver's and Bechtel's examples indeed differ in a crucial respect: some involve a *spatial* while others involve a *temporal* EIO-part-whole relation. Consider the following example:

The change in the conformation of rhodopsin is a stage in the signal transduction pathway. Although the change in rhodopsin is a cause of the electrical signal generated by the cell, scientists do not regard it as a cause of signal transduction. (Craver and Bechtel 2007, 552)

In this example, the part (the change in the conformation of rhodopsin) is an EIO that is a *temporal* EIO-part of the phenomenon-EIO (signal transduction). Craver and Bechtel correctly diagnose, that there cannot be a causal relation between the conformation change of the rhodopsin and signal transduction. In contrast to that, the heart-attack example involves a *spatial* part-whole relation between EIOs. As already mentioned above, Craver and Bechtel argue that the heart attack cannot be considered a cause of the general's death, because the failure of the heart is a part of the "overall mechanism, and the 'effect' is on the whole mechanism". Indeed, this example can be interpreted in two ways, where only one interpretation does not allow for interlevel causation. According to the first interpretation, the part-whole relation involved in this example is a spatial EIO-part-whole relation that holds between the heart's failure and the EIO that is the general's dying, and we assume that the general's dying begins right when the heart begins to malfunction. In this case, the heart's failure and the general's dying

occur in the same space-time region. Hence, there cannot be a causal relation between the two events. According to a second interpretation of this example, interlevel causation is not prohibited. This is the case if we take *the death of the general to be a (the final) temporal EIO-part of the phenomenon* (the dying). In the next section, I will provide a general analysis of how interlevel causation in a mechanism is possible that captures this interpretation of the heart attack example.

5. Interlevel Causation in Mechanisms

Although Craver and Bechtel are right in that there cannot be causal interactions between a phenomenon-EIO and its spatial and temporal EIO-parts, there might still be a way to make sense of interlevel causation in mechanisms. To see that, recall Craver's definition of the mechanistic level relation:

(Levels of Mechanisms) X's ϕ -ing is at a lower mechanistic level than S's ψ -ing if and only if X's ϕ -ing is a component of the mechanism for S's ψ -ing. (Craver 2007, 189)

According to the EIO-interpretation of mechanisms, components of mechanisms are EIOs that are spatial EIO-parts of the phenomenon-EIO (where the spatial EIO-part and the phenomenon are mutually manipulable). Furthermore, in the previous section I have argued that EIOs have temporal EIO-parts. Every phenomenon-EIO S's ψ -ing can be divided into different temporal EIO-parts. The crucial point is that it is plausible to assume that if X's ϕ -ing is at a lower mechanistic level than S's ψ -ing, then X's ϕ -ing is also at a lower mechanistic level *than every temporal EIO-part of S's ψ -ing*.

(Levels of Mechanism)* X's ϕ -ing is at a lower mechanistic level than S's ψ -ing *and every temporal EIO-part of S's ψ -ing* if and only if X's ϕ -ing is a component of the mechanism for S's ψ -ing.

Consider the example provided in Section 2. In this example, the hippocampus's generating spatial maps is at a lower mechanistic level than the mouse's navigating the Morris Water Maze. The mouse's behavior can be divided into different temporal EIO-parts: first, the mouse is put into the pool (the maze), then the mouse starts swimming, it turns left, then it turns right, and so on, until it finds the platform hidden under the water

surface. It is plausible to hold that the hippocampus generating spatial maps is at a lower mechanistic level than all of these temporal EIO-parts.

Why should we accept the modified notion of a mechanistic level? One reason is that denying the modification seems to lead to a contradiction between two claims. On the one hand, it is highly plausible to assume that S's ψ -ing just is the sum of its temporal parts in a particular temporal order (the mouse's navigation behavior is identical to the mouse's turning left, turning right, and so on, until it finds the platform). Hence, if a component X's ϕ -ing is at a lower mechanistic level than S's ψ -ing, and S's ψ -ing = S's ψ -ing's organized temporal parts, one seems to be committed to accepting that X's ϕ -ing is at a lower mechanistic level than the temporal parts of S's ψ -ing. On the other hand, by denying the modification, one would deny that X's ϕ -ing is at a lower mechanistic level than the temporal parts of S's ψ -ing.

A further reason for accepting the modified notion of a mechanistic level is the following: one feature of the notion of a mechanistic level is that it is local—it does not take levels to organize nature or science as a whole but rather only within a mechanism (Craver 2007, 192). Critics have argued that this is a bug rather than a feature (Eronen 2015). Independently of the objections that have been raised against the locality of mechanistic levels in the literature, denying the suggested modification would render the locality of mechanistic levels even stronger and (more) implausible: things would be at different mechanistic levels only if they occurred at exact the same time. To see that: a contester of the modified definition would argue that the original idea motivating the notion of levels of mechanisms was that we get levels from part-whole relations: X's ϕ -ing is at a lower level than S's ψ -ing only if the former is a spatial-EIO part of S's ψ -ing. My modification would deny this: an X's ϕ -ing need not be a spatial-EIO part of any temporal EIO-parts of S's ψ -ing, and still it would be at a lower level than any temporal EIO-part. Now, a contester of the modification could insist that X's ϕ -ing has to be a spatial-EIO part of the temporal EIO-part of S's ψ -ing in order to be at a lower level than this temporal EIO-part. This contester would be committed to the claim that any X's ϕ -ing is at a lower mechanistic level only relative to phenomena that occur *at the very same time*. Hence, mechanistic levels would turn out to have a vertical dimension only.

This is an unwanted consequence for several reasons: first, the notion of a mechanistic level was supposed to capture the idea that phenomena are at higher-levels relative to *all* of the components of their mechanisms. The components of the

mechanisms are taken to be temporally organized in such a way that they give rise to the phenomenon. This usually includes that the components of a mechanism are active only at specific stages of the phenomenon's occurrence. In these cases, it would be strictly speaking false to say that these components are at a lower level than the *whole* phenomenon. Second, given that the relata of the mechanistic level-relation are acting entities that usually operate at different time scales, it is questionable whether temporal synchrony gives rise to an intelligible individuation of the relata. It might turn out that, for example, it is not the long-term potentiation that is at a lower-level than the hippocampus's generating spatial maps but only the second temporal half of it. This would render the notion of a mechanistic level useless with regard to the original aim of making sense of claims such as 'The cell's behavior is at a lower level than the organism's behavior.' Third, if the notion of a scientific level is to make any sense, it has to account for the horizontal dimension as well, i.e. for the idea that things can be at the *same* level. The original definition of a mechanistic level cannot provide this. Craver admits that "there can be no unique answer to the question of when two items are at the same level" (Craver 2007, 192). Based on the EIO-interpretation we can make sense of the same-level relation: two EIOs are at the same mechanistic level iff one is a temporal EIO-part of the other, or they are temporal EIO-parts of the same EIO.

Note that even if (*Levels of Mechanism**) gives up on a strict reading of the parthood condition it still accounts for the idea that a spatial EIO-part (X's ϕ -ing) and the phenomenon (S's ψ -ing) need to be mutually manipulable in order for the former to be at a lower mechanistic level than the latter: for every X's ϕ -ing there has to be a top-down intervention on (a temporal EIO-part) of S's ψ -ing that changes X's ϕ -ing, and there has to be a bottom-up intervention on X's ϕ -ing that changes (a temporal EIO-part of) S's ψ -ing. Being a component in a mechanism for a particular S's ψ -ing does not require the component to be top-down manipulable with respect to *every* temporal EIO-part of S's ψ -ing; neither does it require every temporal EIO-part to be manipulable through X's ϕ -ing. For an X's ϕ -ing to be a component of the mechanism for S's ψ -ing it suffices that there is *at least one* temporal EIO-part of S's ψ -ing that can be changed by changing X's ϕ -ing, and that there is *at least one* temporal EIO-part of S's ψ -ing that induces changes in X's ϕ -ing. If one denies that S's ψ -ing can be changed by changing a temporal EIO-part of it (although denying this seems to me to be rather implausible), the two conditions I present can be regarded as a suggestion of how to make sense of mutual manipulability: mutual manipulability does not hold between a phenomenon as a

whole and a component but rather between temporal EIO-parts of the phenomenon and the component.

Now, how does this modified notion of a mechanistic level render interlevel causation possible? Interlevel causation becomes possible because some spatial EIO-parts (components of the mechanism) of the phenomenon and some of its proper temporal EIO-parts are not themselves related as parts and wholes. Still, according to the modified notion of a mechanistic level, they are at different levels. To get the idea, consider the Figure 5.

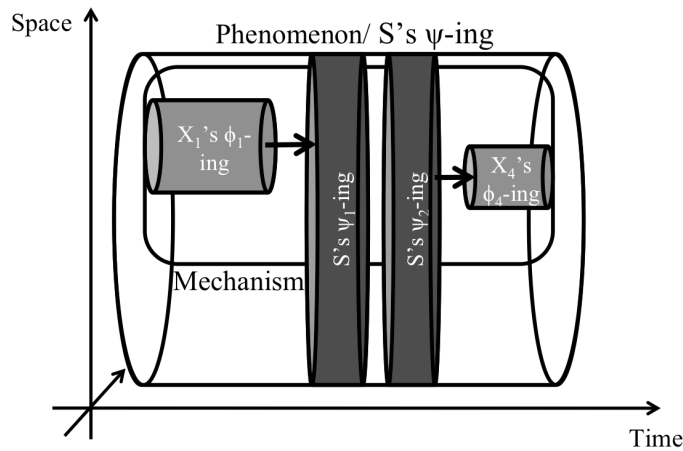


Figure 5 Possible interlevel causal interactions in mechanisms. The X_1 's ϕ_1 -ings are at lower mechanistic levels than S 's ψ -ing and the temporal EIO-parts S 's ψ_j -ings.

In Figure 5, bottom-up as well as top-down causation is illustrated. The *bottom-up* direction consists in an interlevel causal relation between X_1 's ϕ_1 -ing and S 's ψ_1 -ing. On the assumption that X_1 's ϕ_1 -ing is a component of the mechanism for S 's ψ_1 -ing, and our modified definition of a level of mechanism, X_1 's ϕ_1 -ing is at a lower mechanistic level than S 's ψ_1 -ing since S 's ψ_1 -ing is a temporal EIO-part of S 's ψ -ing. Interlevel causation is possible in this case because X_1 's ϕ_1 -ing and S 's ψ_1 -ing are not related as part and whole. Top-down causation, in Figure 5, is illustrated as a relation between S 's ψ_2 -ing and X_4 's ϕ_4 -ing. Similar to the bottom-up direction, in this case S 's ψ_2 -ing and X_4 's ϕ_4 -ing are at different levels. Still, a causal relation between them is possible because they are not related as part and whole. Hence, if we take mechanistic interlevel causation to consist in causal interactions between the components of the mechanism and temporal EIO-parts of the phenomenon, we can make sense of interlevel causation in mechanisms.

Note that I do not want to claim that scientists are usually interested in *explaining* the temporal EIO-parts of a phenomenon. In constitutive mechanistic explanations the

explanandum is a certain phenomenon as a whole—not only temporal EIO-parts of it. Although this might sometimes be the case (for example, when explaining the propagation of an action potential, scientists are interested in how the different phases (i.e. temporal EIO-parts) of depolarization, hyperpolarization, etc. are brought about), this is not what I am arguing for. Rather, my goal in this paper was to provide a metaphysical analysis of the notion of interlevel causation in the context of the new mechanistic approach. Furthermore, my goal is to show that sentences like “her heart attack caused her death” can be literally true. If we take the death to be a temporal EIO-part of the process of dying (plausibly it is the last temporal EIO-part), the heart attack can be a cause of the death in a literal sense even if the death occurs on a higher mechanistic level than the heart attack.

6. Conclusion

Craver and Bechtel’s objection against interlevel causation, summarized in what I called the *Metaphysical Argument*, is valid (i) if one takes the relata to be EIOs (events), and (ii) if one takes lower-level EIO to be spatial EIO-parts of higher-level EIOs. Still, the mechanistic picture leaves room for interlevel causation if one accepts a slight and highly plausible modification of the notion of a mechanistic level: a component of a mechanism is at a lower mechanistic level than the phenomenon *and all of its temporal EIO-parts*. If we take mechanistic interlevel causation to hold between components of mechanisms and temporal EIO-parts of the phenomenon, interlevel causation in mechanisms becomes possible. This view of mechanistic interlevel causation has several advantages: first, we can account for the literal truth of causal statements such as “the heart attack caused her death”; second it makes sense of the idea that levels of nature have a horizontal dimension, and thus it can make sense of the idea that things can be at the same level of mechanism (if one is a temporal part of the other, or if they are temporal parts of the same EIO); third, it might provide resources for a new approach to mental causation. Spelling out the latter must be the topic of another paper.

Acknowledgements

I am thankful to Daniel Brooks, Alexander Dinges and Felipe Romero for comments on earlier versions of the paper. Furthermore, I thank the audience of the Carnap Lectures 2015 in Bochum and Albert Newen and my colleagues at RUB for helpful discussions

on that topic. Finally, I want to thank four anonymous referees for their constructive feedback.

References

- Baumgartner, Michael, and Alexander Gebharter. 2015. "Constitutive Relevance, Mutual Manipulability, and Fat-Handedness." *British Journal for the Philosophy of Science* 67 (3): 731–56. doi:10.1093/bjps/axv003.
- Bechtel, William. 2008. *Mental Mechanisms. Philosophical Perspectives on Cognitive Neuroscience*. New York/London: Routledge.
- Casati, Roberto, and Achille Varzi. 2015. "Events." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta, Winter 201. Metaphysics Research Lab, Stanford University.
- Casini, Lorenzo, and Michael Baumgartner. 2016. "An Abductive Theory of Constitution." *Philosophy of Science*. doi:10.1086/690716.
- Craver, Carl F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York: Oxford University Press.
- . 2014. "The Ontic Account of Scientific Explanation." In *Explanation in the Special Sciences: The Case of Biology and History*, edited by Marie I Kaiser, Oliver R Scholz, Daniel Plenge, and Andreas Hüttemann, 27–52. Dordrecht: Springer Netherlands. doi:10.1007/978-94-007-7563-3_2.
- Craver, Carl F., and William Bechtel. 2007. "Top-down Causation Without Top-down Causes." *Biology & Philosophy* 22 (4): 547–63. doi:10.1007/s10539-006-9028-8.
- Eronen, Markus I. 2015. "Levels of Organization: A Deflationary Account." *Biology and Philosophy* 30 (1): 39–58. doi:10.1007/s10539-014-9461-z.
- Gebharter, Alexander. 2015. "Causal Exclusion and Causal Bayes Nets." *Philosophy and Phenomenological Research*, 1–23. doi:10.1111/phpr.12247.
- Gillett, Carl. 2010. "Moving beyond the Subset Model of Realization: The Problem of Qualitative Distinctness in the Metaphysics of Science." *Synthese* 177 (2): 165–92. doi:10.1007/s11229-010-9840-1.
- . 2013. "Constitution, and Multiple Constitution, in the Sciences: Using the Neuron to Construct a Starting Framework." *Minds and Machines* 23 (3): 309–37. doi:10.1007/s11023-013-9311-9.
- Glennan, Stuart. 2010. "Mechanisms, Causes, and the Layered Model of the World." *Philosophy and Phenomenological Research* 81 (2): 362–81. doi:10.1111/j.1933-1592.2010.00375.x.
- Illari, Phyllis McKay, and Jon Williamson. 2013. "In Defence of Activities." *Journal for General Philosophy of Science* 44 (1): 69–83. doi:10.1007/s10838-013-9217-5.
- Kim, Jaegwon. 2005. *Physicalism, or Something near Enough*. Princeton: Princeton University Press.

- Krickel, Beate. (Forthcoming). "Saving the Mutual Manipulability Account of Constitutive Relevance." *Erkenntnis*.
- . (Under Review). *The Metaphysics of Mechanisms*.
- Leuridan, Bert. 2012. "Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms." *British Journal for the Philosophy of Science* 63 (2): 399–427. doi:10.1093/bjps/axr036.
- Lewis, David K. 1973. "Causation." *Journal of Philosophy* 70 (17): 556–67. doi:10.2307/2025310.
- . 1986. "Events." In *Philosophical Papers Vol. II*, edited by David Lewis, 241–69. Oxford University Press.
- . 2000. "Causation as Influence." *The Journal of Philosophy* 97 (4). JSTOR: 182–197.
- Machamer, Peter, Lindely Darden, and Carl F. Craver. 2000. "Thinking About Mechanisms." *Philosophy of Science* 67 (1): 1–25.
- Psillos, Stathis. 2004. "A Glimpse of the Secret Connexion: Harmonizing Mechanisms with Counterfactuals." *Perspectives on Science* 12 (3): 288–319. doi:10.1162/1063614042795426.
- Quine, Willard V. 1948. "On What There Is." *The Review of Metaphysics* 2 (5). Philosophy Education Society Inc.: 21–38. <http://www.jstor.org/stable/20123117>.
- Robb, David, and John Heil. 2014. "Mental Causation." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta, Spring 201. Metaphysics Research Lab, Stanford University.
- Romero, Felipe. 2015. "Why There Isn't Inter-Level Causation in Mechanisms." *Synthese* 192 (11): 3731–55. doi:10.1007/s11229-015-0718-0.
- Simons, Peter, and Joseph Melia. 2000. "Continuants and Occurrents." *Proceedings of the Aristotelian Society, Supplementary Volumes* 74. [Aristotelian Society, Wiley]: 59–92. <http://www.jstor.org/stable/4107052>.
- Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.
- . 2015. "Interventionism and Causal Exclusion." *Philosophy and Phenomenological Research* 91 (2): 303–47. doi:10.1111/phpr.12095.