

# Constitutive Relevance & Mutual Manipulability Revisited

Carl F. Craver, Stuart Glennan, and Mark Povich

Forthcoming in *Synthese*

Word count: 10558

## **Abstract**

An adequate understanding of the ubiquitous practice of mechanistic explanation requires an account of what Craver (2007) termed “constitutive relevance.” Entities or activities are constitutively relevant to a phenomenon when they are parts of the mechanism responsible for that phenomenon. Craver’s mutual manipulability (MM) account extended Woodward’s account of manipulationist counterfactuals to analyze how interlevel experiments establish constitutive relevance. Critics of MM (e.g., Baumgartner and Gebharder 2016 and Baumgartner and Casini 2017) argue that applying Woodward’s account to this philosophical problem conflates causation and constitution, thus rendering the account incoherent. These criticisms, we argue, arise from failing to distinguish the semantic, epistemic, and metaphysical aspects of the problem of constitutive relevance. In distinguishing these aspects of the problem and responding to these critics accordingly, we amend MM into a refined epistemic criterion, the “matched interlevel experiments” (MIE) account. Further, we explain how this epistemological thesis is grounded in the plausible metaphysical thesis that constitutive relevance is causal betweenness.

## 1. Introduction

In many areas of science, knowledge of mechanisms is prized for its distinctive explanatory power. This knowledge is typically organized across multiple “levels,” in the sense that it includes knowledge of things in different size ranges, knowledge from many domains of science, and, most fundamentally, knowledge of things arranged at least approximately in iterated part-whole relations. Mechanisms have parts, which themselves have parts, and so on. That we organize our mechanistic knowledge into these hierarchical structures is a byproduct of our most basic explanatory commitment— to understand how things work in terms of the organization and interaction of their parts.

Think, for example, of what we know about locomotion in the roundworm nematode, *C. elegans*. We know all its neurons and their connections both to one another and to the muscles (organs, tissues) they innervate. This anatomical and physiological background knowledge provides active, spatial, and structural constraints on any possible mechanism this machinery implements (see, e.g., Sengupta and Samuel 2009). For example, neuroscientists now understand, start to finish, how worms back up when you touch them on the “head”. This explanation spans multiple levels of organization: It includes facts about worms (e.g., *C. elegans*), muscles (e.g., lining the anterior ventral body), neuronal networks (e.g., producing sinusoidal rhythms), neurons (e.g., ALML and ALMR sensory neurons for head touch and the VA and DA motor neurons for control of motion), receptors (e.g., glutamate receptors), and ions (e.g., sodium ions). The touch receptors on the head must be capable of transforming a mechanical stimulus into an electrical signal, and this signal must be communicated from the ALML and ALMR neurons, via intervening networks, to the VA and DA motor neurons and

muscles; that transduction occurs via neurotransmitters, such as glutamate. In cases like this, a behavior is explained in terms of biological mechanisms and ultimately the activities of the molecular and ionic parts of which these mechanisms are composed.

The scientific demand for multilevel mechanistic explanation calls out for philosophical analysis and clarification: What is this multilevel, mechanistic world-picture? How does it shape both the structure of our explanations and the methods by which they are discovered? In his pioneering *Scientific Explanation and the Causal Structure of the World*, Salmon (1984) characterized causal-mechanical explanation as a matter of knowing how an explanandum event, E, is situated in the causal structure of the world. He distinguishes two aspects of causal-mechanical explanation: etiological and constitutive. The etiological aspect looks back and reveals E's antecedent causes. Such aspects are emphasized in, for example, explaining the evolution and development of the worm. The constitutive aspect looks within E to identify its components and their organization. In worm physiology, for example, one looks within the worm and shows how its parts are organized and interact with one another such that they exhibit the phenomenon.

Salmon's remarks on the constitutive aspect are few and brief. Indeed, the topic was largely ignored until philosophers began paying more attention to mechanistic discovery and explanation in the special sciences (Bechtel and Richardson 1993; Craver 2001; Machamer, Darden and Craver [MDC] 2000; Darden and Craver 2001; Glennan 1996, 2002). Craver (2001; 2005; 2007) argued that this "new mechanism" could flesh out the constitutive aspect of Salmon's analysis. At the heart of Craver's analysis is his "mutual manipulability" (MM) account of constitutive relevance. Craver argued that philosophers should be guided in their thinking

about constitutive relevance by attending to the *interlevel experiments* scientists use to test whether something is a component in a mechanism. MM adapts Woodward's (2003) manipulationist analysis of causal relevance to characterize relationships between variables describing things at different levels of mechanistic organization. The promissory notes in Salmon's work, Craver argued, would be completed by adapting this account of relevance for application to part-whole relations in addition to cause-effect relations.

The importance of constitutive explanations is now widely recognized and uncontroversial, but controversy persists over whether Craver's MM account succeeds. Some argue that MM applies the manipulationist approach to a domain that prohibits its application, rendering the account incoherent, absurd, and chimerical (Baumgartner and Gebharder 2016 and Baumgartner and Casini 2017, henceforth BGC<sup>1</sup>; see also Harbecke 2010, Leuridan 2012, Harinen 2018). Specifically, BGC argue that MM requires ideal interventions on wholes with respect to their parts, which is conceptually confused. These critics identify some key obstacles in the way of a coherent and systematic understanding of both the metaphysics and the epistemology of constitutive relevance and interlevel experiments. Here, we remove these obstacles by providing a novel "matched interlevel experiments" (MIE) criterion of constitutive relevance that retains the spirit of MM without conceptual confusion.

---

<sup>1</sup> We treat Baumgartner, Gebharder, and Casini's as a single critical commentary, though we acknowledge that they might not all agree on every point.

Philosophical progress on this problem has been hampered by a failure to disentangle three sorts of question. First are the *conceptual* questions, such as: What do terms like “mechanism” and “component” mean? Does the term “constitution” refer to one kind of relation or many? Second, there are *epistemological* and *methodological* questions: What would count as sufficient evidence, in practice, to establish a component’s constitutive relevance? Why is this evidence convincing? Last are the *metaphysical* questions: What features of the world make a component constitutively relevant to a mechanism?

Craver formulated MM as an answer to the epistemological question. Perhaps because Craver bracketed matters of metaphysical interpretation, many have mistakenly construed MM as an ontological thesis. This interpretive shift is one source of the resulting confusion. We remove this confusion by fleshing out the commitments of MM (revising it into MIE), making its epistemic character more explicit, and offering an ontological theory of constitutive relevance that explains plausibly why MIE is evidentially sufficient.

We proceed as follows: In section 2, we address core conceptual questions about what mechanisms and their parts are and how mechanisms are organized in levels. In section 3, we review Craver’s MM account and explore why critics have claimed it is incoherent. In section 4, we reinterpret interlevel experiments in a way that dissolves this apparent incoherence. Specifically, we characterize more precisely the interventions involved in interlevel experiments and add a matching condition. This is the suitably improved MIE criterion, which we offer as a sufficient, epistemic condition for establishing constitutive relevance. In section 5, we consider the ontological import of MIE, arguing that our methodology points to “causal betweenness” as the truthmaker for claims about constitutive relevance. Given that MIE is offered only as a

sufficient condition, we consider in section 6 conditions under which MIE might fail to detect constitutive relevance, and we list some other kinds of evidence that might be used in those contexts.

## 2. Background: Mechanisms and Levels

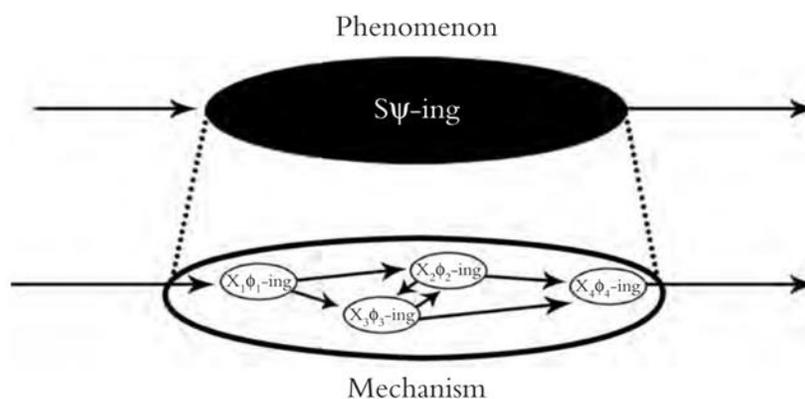
The puzzles of constitutive relevance arise in the context of building a theory of scientific explanation capable of accommodating the multilevel explanations found in sciences such as cell biology, ecology, and solar physics. To give a *causal* (Salmon says “etiological”) explanation for an explanandum event, E, one must determine which among the objects, processes, and events in E’s past are causally (and so explanatorily) relevant to its occurrence. In contrast, to give a constitutive explanation for E involves determining which objects, processes, and events constitute E: some entities and activities within the worm contribute to its turning when touched, and some do not. Both etiological and constitutive explanations require a corresponding notion of “relevance” to sort the things that belong in the explanation from those that do not.

We take a mechanism to be a collection of entities whose activities and interactions are organized such that they are responsible for some phenomenon.<sup>2</sup> The worm’s response to a

---

<sup>2</sup> As Illari and Williamson (2012) note, the ontological significance of the many formulations of “mechanism” is often overstated. We use Glennan’s (2017) definition of “minimal mechanism”, which itself borrows MDC’s 2000 convention of distinguishing “entities,” the component parts, and “activities,”

touch on its head is a phenomenon, and the mechanism responsible for this phenomenon consists of various entities within the worm (e.g., muscle cells and neurons) acting and interacting (e.g., contracting and signaling) in coordinated ways. Those entities, activities, and interactions are the parts or components of the mechanism. Constitutive relevance is the relation between relevant parts (entities and activities) and the mechanism as a whole.



**Figure 1: A Phenomenon and its Mechanism, redrawn from Craver 2007**

Figure 1 replicates a now-standard representation of a mechanism. It shows the relationship between a phenomenon and its mechanism. Roman letters (S and X<sub>i</sub>) represent entities; Greek letters (φ<sub>i</sub> and ψ) represent activities. The figure shows the activity ψ of a composite entity S as made up of the organized φ<sub>i</sub>-ing of X<sub>s</sub>. The arrows indicate at once spatial, temporal, and causal relations among these. In our example, the worm (S) turns-when-touched

---

the component processes. Interactions are activities among two or more entities (Tabery 2004). See Glennan 2017, Kaiser 2018, and Krickel 2018.

( $\psi$ ). Some parts of the worm are relevant to that; some aren't. The relevant ones (such as the AVLM neurons) are the mechanism's constituents. The AVLM neuron's generation of an action potential is likewise an activity ( $\phi$ ) of an entity  $X$  (the neuron) that is part of the activity of turning. Looking down another level, we could describe the organized  $\sigma_j$ -ing of  $P_s$  responsible for  $X$ 's  $\phi$ -ing, and so on.

Crucially, all mechanisms are defined relative to their activities ( $\psi$ ), i.e., defined relative to what they do. There is no mechanism of the worm or of its central nervous system. Claims about constitutive relevance are framed with respect to  $\psi$ . If we remove  $\psi$  from Figure 1, we no longer have a mechanism. Similarly, if we remove all the  $\phi_i$ , the would-be mechanism would be like a stopped clock. If we say a stopped clock has parts, it is because we first think of it running, and then determine which parts are relevant to that. For this conceptual reason, a temporally organized process lies at the core of every mechanism.<sup>3</sup>

These observations reveal two misleading aspects of Figure 1. First, the crisp boundaries of the circle at the top (and in the mechanism beneath it) suggest that  $S$  (and the  $X_s$  that compose it) must be well-defined, localized entities, and that what it is to be mechanism is to be an "entity acting" (an  $S$  that is  $\psi$ -ing). But the boundaries of mechanisms and the boundaries

---

<sup>3</sup> One persistent criticism of the mechanistic approach is that it ignores the processual character of biological systems (Nicholson 2012; Dupré 2013; Dupré and Nicholson 2018). This is puzzling given that processes were at the heart of Salmon's philosophy and that more recent mechanistic approaches have often adopted that processual orientation explicitly (see, e.g., MDC 2000, Craver 1998; Glennan 2017).

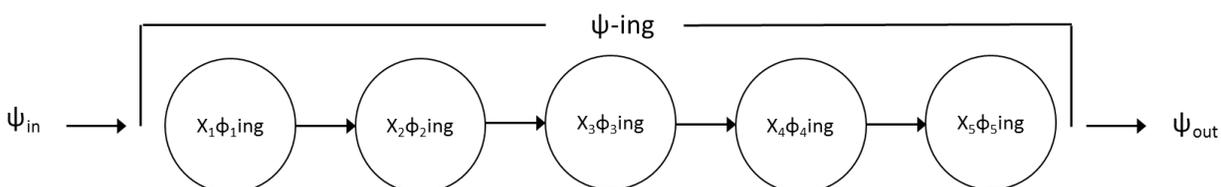
of entities like bodies, cells, and nuclei often do not coincide. Many mechanisms are mechanisms by which two or more entities interact, as in the mechanisms of synaptic transmission. And some mechanisms, like erosion on a riverbank or the Rayleigh scattering that makes the sky blue, are not embodied in entities; they are not mechanisms by which an entity acts, or by which a collection of entities interact. In such cases there is no determinate entity  $S$  that  $\psi$ s, or an  $S$  and a  $T$  that  $\psi$  with each other; there is only  $\psi$ -ing.<sup>4</sup> Second, and more beguiling, Figure 1 combines spatial and synchronous relations of part-whole relations among entities (i.e., between  $X$  and  $S$ ) with the temporal part-whole relations among mechanistic processes at different levels. One consequence is that the bidirectional arrows representing interactions become metaphysically perplexing, as they represent cycling backward and forward in time and causal order (cf. Gebharter 2017; Leuridan 2012; Kim 2000).

To avoid these implications, we replace this diagram with another that maintains all the key relations but removes these misleading implications (Figure 2). Here,  $\psi$ -ing is represented as a process beginning with an input,  $\psi_{in}$ , and terminating with an output,  $\psi_{out}$ . Between these temporal endpoints, and a mechanistic level down, is a temporally sequenced causal chain of events, involving the  $X_i$  and their various  $\phi_i$ . Figure 2 makes clear that the problem of constitutive relevance is that of identifying the components of the process bridging  $\psi_{in}$  and  $\psi_{out}$ : What lies on the causal path(s) between these phenomenon-defining endpoints? The

---

<sup>4</sup> Reflection on examples suggests that embodiment (in an  $S$ ) comes in degrees. But to elucidate and defend this claim would require us to articulate more clearly the criteria by which entities are identified and individuated, and the conditions under which a collection of entities compose another entity. See Gillett 2016 and Glennan 2020 for recent discussions. Because the constitutive relevance problem is about the processual core of mechanisms, we bracket this question.

higher-level activity,  $\psi$ -ing, just is an organized collection of  $X_i \phi_i$ -ing. When we touch the worm on the head it reverses. By what intermediate steps is the tap on the head transduced into a backward movement? We answer this question by describing a mechanism.



**Figure 2. The Processual Core of a Mechanism.**

This processual diagram is also potentially misleading, albeit in different ways. As mechanists have emphasized from the start (see MDC 2000),  $\psi_{in}$  and  $\psi_{out}$  are themselves frequently complex, involving states of many entities and their activities. And the stages between them might more aptly be visualized as a rope of densely entwined fibers than as a chain of separable events. The process might branch and join multiply and have built-in redundancies. The activities of the components need not involve single components but might also involve interactions among multiple components; and the same, particular components can appear time and again at different stages (as the lungs contribute to our breathing or our hearts to the circulation of our blood). Nothing in the diagram is intended to rule out feedback relations and causal cycles, though we hold that such relations, properly understood, can be represented, and should be understood, sequentially in time (see, e.g., Gebharter 2017). Despite these potential limitations, Figure 2 is particularly useful in articulating the (processual) sense of constitution that underlies the logic of interlevel experiments. Though it is possible to translate back and forth between Figure 1 and Figure 2, Figure 2 highlights more clearly the

nature of the interlevel relation, and thus provides a better starting place for understanding MM (and its successor, MIE).

The mechanism by which the worm turns in response to a head-tap spans many levels. The whole worm's reversal depends on neurons firing and muscles twitching, which in turn rely on channels gating and charged particles diffusing, and so on. Following Craver (2007), we take levels of mechanisms to be defined in terms of constitutive relevance relations: if a mechanism is responsible for a phenomenon, the entities, activities, and interactions constitutively relevant to that phenomenon are at a lower level than the mechanism behaving as a whole. Figure 1 represents two levels for an embodied  $S$ :  $S$ 's  $\psi$ -ing is made up of  $X$ s and their  $\phi_i$ -ings. The  $X$ s in this case are all parts of  $S$ ; further, they are  $S$ 's relevant parts— the parts in virtue of which  $S$   $\psi$ s.  $S$ 's  $\psi$ -ing is also partly constituted by the  $\phi_i$ -ings.

Figure 2, which now replaces Figure 1, also can be described in terms of levels, even though it dispenses with the top-level entity  $S$  in Figure 1. There is a level of input-output relations (i.e., a “phenomenal” level), and there is a level of the component parts and activities in virtue of which that input-output relationship is maintained. In Figure 2,  $\psi$ -ing is the higher-level phenomenon; it is constituted by the organized  $\phi_i$ -ing of  $X$ s.  $\psi_{in}$  is the input to the mechanism (the tap on the head) and  $\psi_{out}$  is understood as a downstream consequence of its operation (the backward movement). Because  $\psi_{in}$  and  $\psi_{out}$  lie outside the process, these variables should not be understood as being at a “higher-level” or “lower-level” than the mechanism they bound. As Craver (2007) argues, mechanistic levels are defined only within local hierarchies of mechanistic part to whole. While some might take Figure 2 to do away with levels, it does not. Properly understood, “mechanistic level” refers to what stands in the

abovementioned constitutive relations:  $\psi$ -ing is therefore at a higher level than the various  $\phi$ -ings (cf. Wilhelm 2019). The spatial metaphor of above and below can sometimes make this relation sound much more mysterious than it in fact is.

This conception of mechanistic levels, as mechanists and their critics agree, precludes causal interactions between things at different mechanistic levels. Such interactions are, as Lewis argues, impossible because causes must be distinct from their effects:

If C causes 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 non-implication. 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. (Lewis 2000, p. 78)

Since entities engaging in activities are events, we may apply this principle to a mechanism. The  $\phi$ -ings that are constitutively relevant to a mechanism are parts of, and hence at a lower level than, the  $\psi$ -ing they constitute, and thus cannot cause or be caused by the  $\psi$ -ing. The generation of an action potential in an ALML sensory neuron, for example, does not cause the rising phase of that very action potential; rather, that token action potential is partly constituted by its rising phase. The would-be cause in this top-down causal claim contains the would-be effect within it.<sup>5</sup>

---

<sup>5</sup> Woodward (forthcoming) notes that the value of the cause variable cannot in this case be fixed independently of the value of the would-be effect variable. This is true, however, precisely for the deeper reason that the two stand in a part-whole relationship to one another, as mechanists emphasize.

Bottom-up causation fares no better. The rising phase of the action potential surely causes this action potential to reach its peak and subsequently fall. But the action potential includes its own rising phase, and the rising phase does not cause itself. This is a conceptual point and not an empirical regularity or a matter of natural law. The putative low-level effect is a precondition for the existence of its putative cause. For  $X_i$ 's  $\phi_i$ -ing to cause  $\psi$ , it would have to, *per impossibile*, cause itself. Our processual representation of levels thus makes clear why the relation between levels of mechanisms cannot be causal:  $\psi$ -ing cannot cause  $\phi$ -ing because  $\phi$ -ing is part of  $\psi$ -ing. This explains both why Craver and Bechtel (2005) reject interlevel causation and why they translate such talk into a "hybrid" conjunction of constitutive and causal claims.<sup>6</sup>

### 3. Mutual Manipulability and the Epistemology of Interlevel Experiments

With these conceptual anchors in place, we can now ask: How do scientists determine whether an entity (e.g., an ALML neuron) or activity (e.g., its action potential) is part of a mechanism? Craver's (2007) mutual manipulability account attempted to answer this question by appealing to experimental methods scientists use to address it. Craver formulates MM as a

---

<sup>6</sup> To reject interlevel causal relations is consistent with emphasizing the import of higher-level causes in multilevel mechanisms. Mechanists generally (Craver and Bechtel 2007; Craver 2007; Glennan 2010, 2017) defend the existence and explanatory relevance of higher-level causes while rejecting the idea of interlevel causation as incoherent. Even if scientists do not routinely guard these metaphysical distinctions, they must be guarded nonetheless if we are to avoid speaking nonsense.

sufficient condition for *establishing* that an activity  $\phi$  of some entity  $X$  is constitutively relevant to the mechanism for  $S$ 's  $\psi$ -ing<sup>7</sup>. According to MM, the following conditions are sufficient to establish constitutive relevance:

- (i)  $X$  is part of  $S$ ;
- (ii) in the conditions relevant to the request for explanation, there is some change to  $X$ 's  $\phi$ -ing that changes  $S$ 's  $\psi$ -ing; and
- (iii) in the conditions relevant to the request for explanation, there is some change to  $S$ 's  $\psi$ -ing that changes  $X$ 's  $\phi$ -ing.

Conditions (ii) and (iii) are further elaborated using Woodward's (2003) notion of ideal interventions as follows:

- (CR1) When  $\phi$  is set to the value  $\phi_1$  in an ideal intervention, then  $\psi$  takes on the value  $f(\phi_1)$ ; and

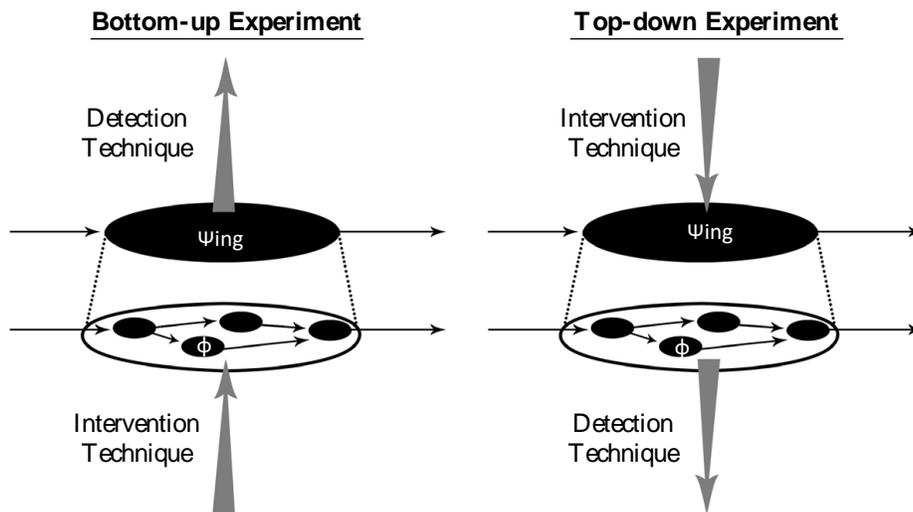
---

<sup>7</sup> Some (e.g., Baumgartner and Casini 2017 and two anonymous reviewers) interpret Craver as offering a necessary and sufficient condition, but the textual evidence against this interpretation seems unambiguous. Craver describes MM as a sufficient condition (2007a, 104, 141, 159 twice; 2007b, 17). His formal specification is articulated using "if" (2007a, 154) or "when" (153; 2007b, 15) rather than "only if," "if and only if," or "when and only when." Finally, he offers counterexamples to show the thesis is not a necessary condition (2007a, beginning on 159; 2007b, 17) and so only a "suitable starting point for an account of constitutive relevance" (2007a, 160).

(CR2) When  $\psi$  is set to the value  $\psi_1$  in an ideal intervention, then  $\phi$  takes on the value  $g(\psi_1)$ .

This epistemic criterion is inspired by three kinds of interlevel experiment commonly used to establish that an entity or activity is relevant to a phenomenon (Craver 2001; see also Stevens 1998; LeDoux 2003): top-down excitatory experiments, bottom-up excitatory experiments, and bottom-up inhibitory experiments. An adequate theory of what constitutive relevance is should help explain why these experiments suffice to detect it.

Craver presented these experiments schematically in Figure 3, building on the representation in Figure 1. On the left is a bottom-up experiment, such as a lesion experiment, in which one intervenes, for example, to remove or inhibit a part (such as the ALML neuron) and then detects the consequences for the behavior of the mechanism as a whole (the failure to reverse). On the right is a top-down excitatory experiment. For example, one taps the worm's head and records from the ALML neuron as the worm reverses; the goal (roughly) is to determine whether the component's activity changes as the mechanism acts.



**Figure 3: Craver (2007) Representation of Bottom-up and Top-down Experiments**

Figure 3, however, is easily misinterpreted. The grey arrows representing interventions and detections appear to jut into and out of the mechanism ambiguously, leaving the relationship between the intervention and the detection unclear. Combined with the Woodwardian view of causal relevance, Figure 3 invites the conclusion that the thing on which we intervene and the thing from which we detect must be causally related. Such experiments thus seem to establish what everyone in the discussion agrees cannot be: namely, causal relations between parts and wholes.

Indeed, BGC believe MM embodies just this mistake. They argue that MM is conceptually incoherent (Baumgartner and Gebharter 2016; Baumgartner and Casini 2017)<sup>8</sup> because there is

---

<sup>8</sup> Baumgartner and Cassini's critique of MM rests in part on treating MM as a necessary and sufficient condition. Their positive argument for treating it as such despite the textual evidence in note 7, we believe, comes from a misreading of MM. They argue as follows:

MM provides a sufficient condition for constitutive relevance and a sufficient condition for constitutive irrelevance, which jointly amount to a sufficient and necessary condition for constitutive relevance—that is, to a complete definition of constitution (Baumgartner and Casini 2017, p. 218).

But their reasoning is fallacious. Consider the following formalization:

R: X's  $\phi$ -ing is constitutively relevant to S's  $\psi$ -ing (constitutive relevance)

B: manipulating  $\phi$  changes  $\psi$  (bottom-up manipulability)

T: manipulating  $\psi$  changes  $\phi$  (top-down manipulability)

Baumgartner and Casini take Craver to assert the following sufficient conditions for relevance and irrelevance respectively:

$$(1) (B \ \& \ T) \supset R$$

$$(2) \sim(B \ \& \ T) \supset \sim R$$

Together, (1) and (2) are indeed equivalent to a necessary and sufficient condition:

$$(3) R \equiv (B \ \& \ T)$$

a *prima facie* mismatch between the requirements of Woodward's theory of causal relevance and the mechanistic understanding of levels. Specifically, restrictions on the former *prevent* their application to the latter.

To see why, note that the counterfactuals in CR1 and CR2 are Woodwardian, manipulationist counterfactuals (Woodward 2003). Woodward's view of causal relevance holds that a variable (X) is causally relevant to another variable (Y) if and only if one can change the value of Y by intervening to change the value of X. This rough statement requires restrictions and caveats. The key restriction is that the intervention (I) on X must not change Y via any route other than X. In particular, I cannot change Y directly. If I changes Y directly, then the observed change in Y might not be due to the change I induces in X, but rather to effects propagated

---

However, Craver does not assert (2), and it appears Baumgartner and Casini have misplaced the negation in (2). Craver's actual statement of the sufficient condition is:

To establish that a component is irrelevant, it is sufficient to show that one cannot manipulate S's  $\psi$ -ing by intervening to change X's  $\phi$ -ing and that one cannot manipulate X's  $\phi$ -ing by manipulating S's  $\psi$ -ing (Craver 2007, 159).

Using our formalization:

$$(2C) (\sim B \ \& \ \sim T) \supset \sim R$$

Crucially, 1 and 2C do not together entail a biconditional. As Craver writes, in cases where B is true but T is false, or vice versa, constitutive relevance is indeterminate.

along some other path. In that case, X may be causally irrelevant to Y even though interventions on X change Y.

For BGC, the incoherence of MM arises from a further premise: that the phenomenon *supervenes* on the causal organization of the mechanism's parts. They argue that there can be no difference in S's  $\psi$ -ing in a context without a difference in the parts, properties, activities, and organizational features of the mechanism (the  $X_i$ 's  $\phi_j$ -ings). Consequently, one cannot intervene to change S's  $\psi$ -ing without changing at least one of the  $X_i$ 's  $\phi_j$ -ing. The supervenience of S's  $\psi$ -ing on the  $X_i$ 's  $\phi_j$ -ing, BGC argue, makes it impossible to intervene ideally on S's  $\psi$ -ing with respect to any of the  $X_i$ 's  $\phi_j$ -ings. Thus, BGC claim, CR2 is incoherent. Given the definition of an ideal intervention, CR2 demands that one must be able to change S's  $\psi$ -ing without also directly or indirectly intervening into  $X_i$ 's  $\phi_j$ -ing; and given that S's token  $\psi$ -ing supervenes on  $X_i$ 's token  $\phi_j$ -ings, that is incoherent.

Baumgartner and Gebharder (2016, Figure 3, p. 742) consider three ways to make sense of the ability to intervene on  $\psi$  ( $I_\psi$ ) with respect to an X and its  $\phi$ -ing, all of which fail:

- (A)  $I_\psi \rightarrow \psi \rightarrow \phi_j$ ; that is, one intervenes on  $\psi$ , which then causes a change to  $\phi_j$ ;
- (B)  $I_\psi \rightarrow \phi_j \rightarrow \psi$ ; that is, one intervenes on  $\phi_j$ , which then causes a change to  $\psi$ ; or
- (C)  $I_\psi$  is a common cause of  $\phi_j$  and  $\psi$

As Baumgartner and Gebharder argue, A and B violate the supposition that there cannot be interlevel causation, and C violates the constraint that an intervention should not change both the putative cause and its effect at the same time.

Though Baumgartner and Gebharder explicate the problem by assuming supervenience, the putative incoherence can be expressed without appeal to that assumption.<sup>9</sup> These three principles are sufficient:

- **Part-Whole Relation:**  $X_i$ 's  $\phi_j$ -ing is part of ( $S$ 's)  $\psi$ -ing.<sup>10</sup>
- **Interlevel Intervention:** CR2 requires interventions on  $\psi$  with respect to  $X_i$ 's  $\phi_j$ -ing.
- **Manipulationism:** If  $X_i$ 's  $\phi_j$ -ing is part of ( $S$ 's)  $\psi$ -ing, then it is not possible to intervene on  $\psi$  with respect to any  $X_i$ 's  $\phi_j$ -ing.

The incoherence follows, as Lewis argued, from the overlap of the spatiotemporal whole and its spatiotemporal parts, as represented in Figures 2 and 4.

An alternate way to formulate the putative incoherence shows how it arises from illicit interlevel causation:

- **Part-Whole Relation:**  $X_i$ 's  $\phi_j$ -ing is part of ( $S$ 's)  $\psi$ -ing.
- **Interlevel Intervention:** CR2 requires interventions on  $\psi$  with respect to  $X_i$ 's  $\phi_j$ -ing.

---

<sup>9</sup> We do not doubt that wholes supervene on the organized collections of their parts, but the nature of this relation is a matter of dispute. We focus on the part-whole relation because it is better understood, and the putative challenge to MM comes directly from the impossibility of causal relations obtaining between parts and wholes.

<sup>10</sup> Parentheses indicate that  $\psi$  may be more or less embodied and so  $S$  is not essential to the parthood relation we consider.

- **Causal Sufficiency:** If one can change Y via an ideal intervention on X, then X is causally relevant to Y.
- **No Interlevel Causation:** Parts and wholes are not causally related.

Given causal sufficiency and the satisfaction of CR2,  $X_i$ 's  $\phi_j$ -ing and  $S$ 's  $\psi$ -ing must be causally related, in violation of the prohibition against interlevel causation. This is how Romero (2015), partly inspired by Leuridan (2012), formulates the problem.<sup>11</sup>

These two arguments for MM's incoherence share two premises: Part-Whole Relation and Interlevel Intervention. Our response and amended view below reject Interlevel Intervention and, correspondingly, correct the implicit view of how interlevel experiments are understood.

---

<sup>11</sup> Some critics reject high-level interventions as "fat-handed", a term borrowed from, e.g., Scheines 2005. A fat-handed intervention on  $\psi$  changes more variables than  $\psi$ , including possibly  $\phi$ . But fat-handedness *per se* is not the problem here. Fat-handed interventions are problematic when the same intervention simultaneously manipulates two variables, each of which independently could be causally relevant *de facto* to the putative effect. In this case, however, the variables are not independent competitors. Rather, the intervention on the putative cause *necessarily* changes the putative effect. That is not *de facto* fat-handedness but rather a metaphysically necessary fat-handedness. Our two reconstructions of the problem here dispense with this obfusatory ambiguity.

#### 4. Matched Interlevel Experiments (MIE) as an Epistemic Sufficient Condition for Constitutive Relevance

So, how are interlevel experiments possible when things at different levels are, by definition, related as mechanistic part to whole? How can one intervene on a part with respect to a whole and vice versa? The key conceptual reorientation is to interpret interlevel experiments as interventions into temporal parts of a mechanistic process as depicted in Figure 2. Here we explain how the different kinds of interlevel experiment do just that, and we show how those experiments, appropriately combined, provide a sufficient epistemic condition for establishing the constitutive relevance of a part to the behavior of a mechanism as a whole. This is our matched interlevel experiments (MIE) criterion.

Consider again the structure of the three kinds of interlevel experiment. Two are bottom-up, and one is top-down:

(1) **Bottom-up Inhibitory Experiments:** Delete or inhibit a component (e.g., the ALML neuron) and detect the behavior of the mechanism as a whole (e.g., the worm's reversal).

(2) **Bottom-up Excitatory Experiments:** Excite a component (e.g., deliver artificial current to the ALML neuron) and detect the behavior of the whole (e.g., the worm's reversal).

(3) **Top-down Excitatory Experiments:** Engage the phenomenon (e.g., tap the worm's head) and detect how the parts change (e.g., the ALML neuron fires) as the phenomenon unfolds (e.g., the worm's reversal).<sup>12</sup>

The *epistemological* task is to understand why these experimental strategies are relevant to assessing whether an entity or activity is part of the mechanism for a phenomenon. How do these experiments work? Why do they justify this inference? Our answers turn on observing how the interventions and detections in these experiments are situated at different stages in the processual core of the mechanism.

Figure 2 depicts a mechanistic process  $\psi$  as a sequence of stages involving Xs and their  $\phi$ -ing. Using this notation, and following the suggestions of Harinen (2018) and Prychitko (2019), we redescribe these three kinds of experiment as follows:

(1A) **Bottom-up Inhibitory Experiments.** Delete or inhibit a component (X's  $\phi$ -ing).

Intervene to establish startup conditions  $\psi_{in}$ . Measure  $\psi_{out}$ . Evaluate thereby whether X and its  $\phi$ -ing are necessary for  $\psi_{in}$  to produce  $\psi_{out}$ .

(2A) **Bottom-up Excitatory.** Intervene to stimulate X to  $\phi$ . Measure  $\psi_{out}$ . Evaluate

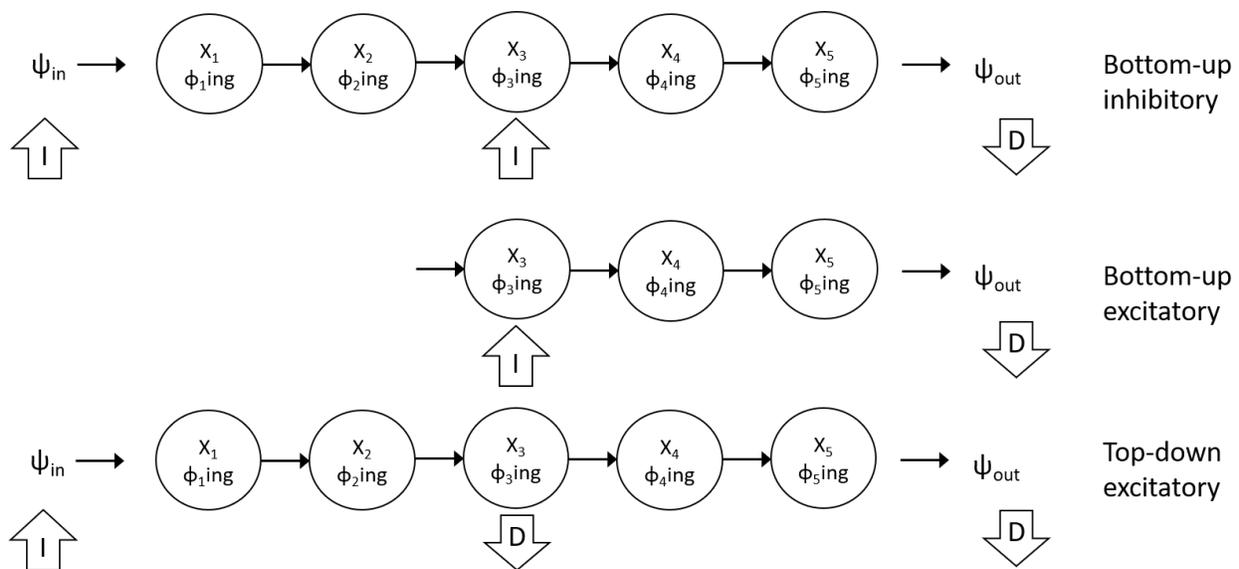
whether one can control  $\psi_{out}$  by manipulating X or its  $\phi$ -ing. Evaluate thereby whether X and its  $\phi$ -ing are causes of  $\psi_{out}$ .

---

<sup>12</sup> Top-down inhibitory experiments are the control conditions for top-down excitatory experiments. Whether an intervention is "excitatory" and "inhibitory" depends on how something is situated in a causal context.

(3A) **Top-down Activation.** Bring about startup conditions  $\psi_{in}$ . Measure  $X$ 's  $\phi$ -ing. Also measure  $\psi_{out}$ . Evaluate thereby whether  $X$  and its  $\phi$ -ing are changed when  $\psi_{in}$  produces  $\psi_{out}$ .

Figure 4, which now replaces Figure 3, represents these experiments in a way that eliminates the visual suggestion that interlevel interventions involve causal relations between parts and wholes.



**Figure 4: A Processual Representation of Interlevel Experiments.**

Using Figure 4, we can reformulate the criteria of mutual manipulability (CR1 and CR2), making the connection with interlevel experiments explicit. We begin by representing the inhibitory and the excitatory interventions from CR1 separately:

(CR1i) If an experiment initiates conditions  $\psi_{in}$  while a bottom-up intervention,  $I$ , prevents or inhibits  $X$ 's  $\phi$ -ing, alterations to or prevention of  $\psi$ 's terminal conditions,  $\psi_{out}$ , are detected.

(CR1e) If a bottom-up intervention,  $I$ , stimulates  $X$ 's  $\phi$ -ing,  $\psi$ 's terminal conditions,  $\psi_{out}$ , are detected.

Many interlevel experiments can be conceptualized as involving Boolean variables with values denoted “**on**” and “**off**”. In such cases, CR1i amounts to the requirement that an experiment that sets  $\psi_{in}$  to **on** and  $X$ 's  $\phi$ -ing to **off** should detect  $\psi_{out}$ 's value to be **off**; and CR1e, amounts to the requirement that an experiment that sets  $X$ 's  $\phi$ -ing to **on** should detect  $\psi_{out}$ 's value to be **on**. Pearl (2012) shows how to generalize this formulation beyond the binary case.

As can be seen in Figure 4, nothing in Woodward's understanding of interventions is violated in either case. Although these experiments are not all of the standard Woodwardian form (“wiggle C, detect E”), they are nonetheless causal experiments (for further discussion of non-Woodwardian but causal experiments, see Craver and Darden 2013). CR1i tests whether  $\phi$  is a necessary link in the causal chain between  $\psi_{in}$  and  $\psi_{out}$ . As Prychitko (2019) and Harinen (2018) emphasize, it tests a relationship between three variables; it asks if  $\phi$ -ing is necessary in order to traverse the causal path between  $\psi_{in}$  and  $\psi_{out}$ . Because the causal transition between  $\psi_{in}$  and  $\psi_{out}$  just is  $\psi$ , we can say that  $\phi$ -ing is necessary for  $\psi$ -ing, but we will thereby elide in one use of the word “necessary” both a causal and a constitutive claim (cf. Craver and Bechtel 2007). CR1e is also a straightforward manipulationist counterfactual to the effect that one can change the output of the mechanism by intervening mid-stream to drive one of the mechanism's components. Neither is even apparently incoherent.

The core of BGC's case for incoherence arose from the supposed incoherence of top-down interventions (CR2). But this problem is revealed to be merely apparent once we rephrase CR2 to make the causal structure of this type of experiment explicit:

(CR2\*) If a top-down experiment initiates conditions  $\psi_{in}$  and detects  $\psi$ 's terminal conditions,  $\psi_{out}$ , X's  $\phi$ -ing is also detected.

As with CR1i and CR1e, we can give these variables a Boolean interpretation, in which case, CR2\* amounts to the requirement that setting  $\psi_{in}$  to **on**, in conditions in which  $\psi_{out}$  is detected as **on**, will result in detecting X's  $\phi$ -ing to be **on**. Prychitko and Harinen note there are three variables in these experiments, not two (as in the standard Woodward counterfactual), but the logic of the experiment is clear. It tests whether  $\psi_{in}$  is a cause of X's  $\phi$ -ing and whether it causes the  $\phi$ -ing in conditions in which  $\psi_{out}$  is **on**.<sup>13</sup> It is commonly said that such experiments are "correlational" because they demonstrate only that  $\phi$ -ing occurs during  $\psi$ -ing. Our reconstruction reveals that this a half-truth. In fact, these experiments ask whether intervening to produce  $\psi_{in}$  drives X's  $\phi$ -ing when it produces  $\psi_{out}$ .

It is now clear why interlevel experiments do not contravene Woodward's constraints on ideal interventions. Each experiment tests a different causal claim. And when those causal claims are conjoined, they add up to sufficient evidence that X and its  $\phi$ -ing lie causally

---

<sup>13</sup> Pearl (2012) generalizes the same idea to non-binary relations. That our reconstruction captures the idea of "causal mediation" developed in the causal modeling literature gives us some confidence that this integration of ontology and epistemology is on the right track. This convergence of qualitative discussions of mechanisms with quantitative work on causal modeling is a desired outcome, not a result to be shunned.

between  $\psi_{in}$  and  $\psi_{out}$ . Constitutive relevance thus amounts to a kind of causal mediation (see Pearl 2012):  $\psi_{in}$  exerts its effect on  $\psi_{out}$  partly or wholly via an antecedent effect on X and its  $\phi$ -ing.

Our reinterpretation of these experiments allows us to state succinctly the misstep in BGC's argument.<sup>14</sup> Their three options (A-C) for understanding interlevel interventions are not exhaustive; we can add a fourth:

(D)  $I_{\psi} \rightarrow \psi_{in} \rightarrow \dots \rightarrow X's \phi\text{-ing} \rightarrow \dots \rightarrow \psi_{out}$

Excitatory experiments test the relationship between  $\psi_{in}$  and  $\phi$  in conditions where  $\psi_{out}$  is produced. Inhibitory experiments show that if you prevent  $\phi$ , you sever the processual bridge between  $\psi_{in}$  and  $\psi_{out}$ . And stimulation experiments show that you can drive  $\psi_{out}$  by stimulating  $\phi$ . Explicated thus, these are clearly experiments for testing "causal betweenness": for showing that X and its  $\phi$ -ing are caused by  $\psi_{in}$ , are causes of  $\psi_{out}$ , and are activated or changed in the process of producing  $\psi_{out}$  from  $\psi_{in}$ .

These revisions of CR1 and CR2 fully address concerns about incoherence, but once the structure of this conjunction is explicit, it's clear that an additional hurdle must be crossed to demonstrate causal betweenness. We cross it by introducing a matching criterion: roughly, the values taken by  $\phi$  in the intervention and detection experiments must match.

---

<sup>14</sup> Note that this reinterpretation is not entirely novel; Craver (2007, 145-6) warned against this confusion and hinted at this solution.

To motivate this matching condition, consider cases where it is violated. The first involves a qualitative mismatch in the kinds of  $\phi$ -ings occurring in the different experiments. Suppose one discovers that when you tap the worm's head with a hair and it retreats, a particular neuron becomes "activated"; for example, it exhibits an oscillating burst pattern. This shows that  $\phi$ -ing ("activation") happens during the transition between  $\psi_{in}$  and  $\psi_{out}$ . Suppose next that an experimenter does a bottom-up excitatory experiment in which she "activates"  $\phi$ -ing, perhaps by delivering a continuous spike train at 80 Hz, and she observes the worm turn. The "activation" produced in the top-down experiment (the oscillation) is different from the "activation" created by the intervention in the bottom-up experiment.  $\phi$  is here used as a "filler term" that ambiguously represents qualitatively distinct activities,  $\phi_{\alpha}$  and  $\phi_{\beta}$ , as if they were the same.

Now consider a quantitative case: interventions on  $\psi_{in}$  change  $\phi$ -ing, and interventions into  $\phi$ -ing make a difference to  $\psi_{out}$ , but the values of  $\phi$ -ing involved in the first causal relation are not the values involved in the second. This can occur if the causes or effects of  $\phi$ -ing are sensitive to magnitude, as in threshold responses or sigmoidal response profiles. Values of  $\phi$  above the threshold trigger responses not triggered beneath the threshold. Consider excitotoxicity as an example. Brushing the worm on the head with a hair causes the ALML neuron to fire faster than it would without the brushing. Hyperstimulating the ALML neuron induces neurotoxicity and can cause the neuron to die. But one cannot kill ALML neurons by brushing the worm on the head with a hair. We can establish experimentally that interventions into  $\psi_{in}$  (the brushing) change  $\phi$ -ing (firing rate), and interventions into  $\phi$ -ing (firing rate)

produce  $\psi_{\text{out}}$  (neurotoxicity), but these causal relationships involve different ranges of  $\phi$ 's values.

The matching requirement thus specifies a qualitative and quantitative matching between interventions and detections:

**(Matching)** The activities  $\phi_j$  activated or inhibited in bottom-up experiments (CR1i and CR1e) must be of the same kind, and occur within quantitatively overlapping ranges with, the activities  $\phi_i$  detected in top-down experiments (CR2\*).

Our formulation of Matching is admittedly vague. We cannot yet succinctly state in a fully general way what overlapping of values or its failure amounts to. Similarly, we appeal to the idea that two activities share a common kind, and we have not provided an account of that. Yet the examples hopefully make the core idea clear: If the activities and quantitative changes generated within bottom-up experiments are not the same kinds and values as the changes detected in top-down experiments, the constitutive relevance of an activity will not have been established.

Putting our four conditions together gives us our matched interlevel experiments (MIE) condition:<sup>15</sup>

---

<sup>15</sup> Note that MIE drops the now-unnecessary parthood condition in Craver (2007). If we take S simply to name the collection of all and only the Xs involved in some  $\psi$ -ing, the parthood condition is tautologous and uninformative, as Leuridan (2012) has urged. This is a side-benefit of our revision.

**(MIE)** To establish that an entity  $X$  and its activity  $\phi$  are constitutively relevant to a mechanism that  $\psi$ s, the following experimental results and matching condition are jointly sufficient:

(CR1i) If an experiment initiates conditions  $\psi_{in}$  while a bottom-up intervention,  $I$ , prevents or inhibits  $X$ 's  $\phi$ -ing, alterations to or prevention of  $\psi$ 's terminal conditions,  $\psi_{out}$ , are detected.

(CR1e) If a bottom-up intervention,  $I$ , stimulates  $X$ 's  $\phi$ -ing,  $\psi$ 's terminal conditions,  $\psi_{out}$  are detected.

(CR2\*) If a top-down experiment initiates conditions  $\psi_{in}$  and detects  $\psi$ 's terminal conditions  $\psi_{out}$ ,  $X$ 's  $\phi$ -ing is also detected.

(Matching) The activities  $\phi$  activated or inhibited in bottom-up experiments (CR1i and CR1e) must be of the same kind and occur within quantitatively overlapping ranges with the activities  $\phi$  detected in top-down experiments (CR2\*).

This revised version of mutual manipulability explains why BGC's objections are off the mark while providing a more detailed understanding of the experiments by which claims of constitutive relevance are evaluated. We now turn to the metaphysical truthmakers of such claims.

## 5. Constitutive Relevance as Causal Betweenness

MM and MIE, again, are epistemic conditions— they aim at justifying the claim that some  $X$  and its  $\phi$ -ing are constitutively relevant to ( $S$ 's)  $\psi$ -ing. Now, following the explication in Part 2, we argue that constitutive relevance is causal betweenness. This ontological thesis explains the epistemic power of MIE.

The somewhat unfamiliar notion of causal betweenness in fact has a long philosophic history. Reichenbach (1956) defined causal betweenness as a three-place probabilistic relation between events.<sup>16</sup> He held (roughly) that an event  $B$  was causally between events  $A$  and  $C$  if i)  $P(C|B) > P(C|A)$  and ii)  $B$  screens off  $A$  from  $C$  (i.e.,  $P(C|B\&A) = P(C|B)$ ). Salmon (1980) argued that this definition failed for reasons common to a host of probabilistic theories of causation. In particular, his objections appeal to the possibility of “making it the hard way” – i.e., producing an outcome (e.g., a goal) via a causal chain that plausibly reduces the probability of the outcome below what it would be had it not occurred (e.g., off the referee’s head). For this reason, Salmon argued that causal betweenness cannot be analyzed exhaustively in terms of

---

<sup>16</sup> Causal betweenness is a three-place relation,  $B(x, y, z)$ , between events, and for some event,  $X_i$ 's  $\phi_i$ -ing, to be constitutively relevant to a mechanism that  $\psi$ s, it must be the case that  $B(\psi_{in}, X_i$ 's  $\phi_i$ -ing,  $\psi_{out})$ . The  $X_i$ 's  $\phi_i$ -ing shown in the lower half of Fig 1, in fact lie causally between  $\psi_{in}$  and  $\psi_{out}$  (as shown in Fig. 2). Although the relata of the betweenness relation are events, we also speak of entities  $X$  or activities  $\phi$  as lying between, when those entities and activities are constituents of the events lie between. We understand events, in this context, as entities engaging in activities or interactions (see Glennan 2017). Betweenness is thus fundamentally a relation between particular events. Further work is required to extend this analysis to types of events.

chance-raising; one must appeal in addition to a continuous causal process linking a distal cause to its effect.

If, as we believe, causal dependence requires a productive process between an effect and its cause (Salmon 1984; Anscombe 1971/1981; MDC 2000; Bogen 2008; Glennan 2017), then for B to lie causally between A and C, there must be a process by which A contributes to the production of B, and a process by which B contributes to the production of C. Exactly what makes one event contribute to the production of another has been subject to considerable debate (e.g., Salmon 1984, 1998, Ehring 1997, Dowe 2000), and much of that literature has understood production in terms of basic physical processes. We hope and believe that further developments (e.g., Glennan 2017) provide resources to understand productivity in a way that allows for productive continuity to hold between higher level-events (like the brushing of the worm's head and the subsequent turning), and to make sense of productive continuity in mechanisms that rely on inhibition, disconnection, and other sorts of dependencies that traditionally trouble physical productivity accounts. While we cannot resolve this question here, we emphasize that, regardless how productivity is ultimately understood, manipulability via experiments (interlevel or otherwise), and with it, the truth of Woodwardian counterfactuals, is evidence for, rather than constitutive of, the productive continuity at the heart of every mechanism.

## 6. Why MIE is Merely Sufficient, not Necessary.

Constitutive relevance, we have argued, is causal betweenness. Interlevel experiments establish causal betweenness. But constitutive relevance is not reducible to MIE. This is because

there are other ways to establish constitutive relevance besides MIE, and there are some kinds of component that MIE cannot detect. Or so we now argue.

Interlevel experiments are not the only way to test for constitutive relevance. One might mine correlations to prune the space of possible mechanisms (e.g., Spirtes, Glymour, and Scheines 2000). One might rely on temporal information about what comes before, after, or temporally between one or more other events. One might also infer facts about causation from spatial and structural facts, such as that  $X_1$  and  $X_2$  are on opposite sides of a membrane, or do not synapse with one another, or are outside one another's light cones. Perhaps we can just see causal relationships sometimes. Just as there are many ways to establish a causal claim, there are many ways to establish a constitutive relevance claim.

More interestingly, however, MIE is blind (as was MM) to certain kinds of mechanistic component. Consider the bottom bracket of a bicycle, which holds the spindle connecting the two arms of the pedal crank system. Let  $\psi$  be the process of translating pushing on pedals ( $\psi_{in}$ ) into forward motion ( $\psi_{out}$ ). To get from  $\psi_{in}$  to  $\psi_{out}$ , the spindle has to rotate around a fixed point in the bike's frame, and to do that, it has to fit in the bottom bracket; furthermore, the bottom bracket must be shaped to allow the spindle to rotate. Rotation ( $\phi$ ) of the spindle ( $X$ ) lies causally between the pushing on the pedal and the moving of the bike; and we could use interlevel experiments to prove it. We could, for instance, intervene to turn the spindle (and the attached chain ring) directly (a bottom-up activation experiment), as one step in meeting MIE.

But could we do this with the bottom bracket? It is clearly part of the bike's drive train; if you remove it or alter its shape, pedaling will get you nowhere (an inhibitory experiment). Still, we cannot drive the bike by intervening on the shape of the bottom bracket the way we can with the spindle. We cannot "excite" a bottom bracket because the bottom bracket is not active. It is causally between  $\psi_{in}$  and  $\psi_{out}$ , because it is part of an event between the pedaling and the moving. But we cannot drive the bike (down a flat road) by changing the bracket. The bracket's presence, like the entire frame, is a stable structure that makes  $\psi$ -ing possible. This toy example is replicated in every mechanism we know: they require (relatively) stable structures as standing conditions to work. In the worm, the neurotransmitter's structure is not changed by the action potential, and the ALML neuron's existence does not depend on touching the worm's head. Such components will not register in top-down experiments precisely because the mechanism depends on their *not* changing with the causal input.

MIE will also often fail to establish constitutive relevance in redundant mechanistic processes. For example, humans have two kidneys, each one of which can regulate plasma osmolality on its own. Usually, both work together; they are actively redundant. If we remove one, the other takes up the slack. An inhibitory experiment that removed a kidney and looked for a change in plasma osmolality would not register an effect (unless we really tax the system). Redundancies are known challenges for difference-making accounts of causation: removing these relevant components makes no difference to the effect if a backup is available.

The fact that constitutive relevance is causal betweenness helps us to understand both why MIE (and MM, its ancestor) is a sufficient condition (because MIE's conditions are enough to identify a part in a causal process between  $\psi_{in}$  and  $\psi_{out}$ ) and why it is *only* a sufficient

condition. This is both because there are many ways to test causal and constitutive claims besides MIE and because the entities and activities involved in the  $\psi$ -ing might be redundant or, alternatively, standing conditions rather than active participants. The ontology of constitutive relevance as causal betweenness thus provides a satisfying and systematic explanation of why the epistemology of interlevel experiments works and fails when it does.

## 7. Conclusion: Next Steps for Constitutive Relevance

To understand how the worm turns, we must appeal to two kinds of relation— relations of cause and effect, and relations of mechanistic part to whole. Emphasis on this latter, constitutive, relation has been a major fruit of the recent emphasis on multilevel mechanistic explanation in the philosophy of science. Craver's (2007) MM was the first effort to establish an explicit criterion for constitutive relevance. Subsequent explorations and critiques of this account have identified difficulties and suggest improved ways of understanding constitutive relevance and its relation to causal relevance. One key lesson from this decade or more of discussion is that philosophical analysis must keep conceptual, epistemic, and ontological questions about mechanisms and constitutive relevance distinct while recognizing that they are systematically connected to one another. This is why the epistemology of interlevel experiments can inform work in the ontology of multilevel mechanisms.

Reflection on these lessons has allowed us make progress on several fronts:

- 1) Conceptually, we have clarified that all mechanisms are processual at their core and that constitutive relevance can be construed as a relationship between a mechanistic process and its parts.

- 2) Relatedly, we have clarified how mechanistic levels can be understood processually, and how higher-level processes can be composed of lower-level constituents.
- 3) We have shown how interlevel experiments can be interpreted as unproblematic causal interventions into different stages of a mechanistic process.
- 4) We have shown how MM, interpreted as an epistemic criterion that appeals to such experiments, involves no conceptual incoherence.
- 5) We have created a revised epistemic criterion, Matched Interlevel Experiments (MIE), which clarifies the nature of the required experiments and adds an important matching condition.
- 6) We have argued that the constitutive relevance relation MIE detects is a three-place relation of causal betweenness, the ontological truthmaker for claims of constitutive relevance.
- 7) We have amplified the case that MIE is epistemic, and only sufficient, by exploring other ways constitutively relevant components can be detected or fail to be detectable by MIE and by providing an alternative, processual ontology of causal betweenness as its ontological target.

Still, much work remains to be done.

First, our emphasis on the proprietary sense of constitutive relevance and mechanistic levels should be taken as an invitation to explore and understand the importance of other kinds of composition. We concur with Aizawa and Gillett (2016, 2019) and Ladyman and Ross (2007) that composition and compositional explanation are a genus of which mechanistic constitution

and constitutive explanation are a species. Further exploration of related species, such as object constitution, will advance a common enterprise.

Second, our account of constitution is developed for tokens, not types. It will be important to extend this analysis to mechanism types, and to extend the epistemological methods discussed here to the testing of type-level or otherwise generic claims. A particularly promising avenue for understanding these relations begins with Mackie's notion of an INUS condition. Couch (2011) and Harbecke (2010) argue that this analysis of causal conditions can be used as criterion for constitutive relevance: mechanistic components are insufficient but necessary parts of collections that are jointly sufficient to produce a mechanistic phenomenon. There is something appealing in this idea, which connects to intuitions about multiple realization and may help to deal with the redundant mechanisms discussed above. But work remains to be done to show how the INUS account can be melded with the epistemic and ontological analyses of token cases offered here.

Third, our discussion of causal betweenness brings to light the importance of productive continuity in processes. More work should be done exploring both the sense of productivity involved and how productive processes depend upon their constitutively relevant parts. Whatever productivity is, it clearly must allow for the ubiquity of mechanisms in the life sciences and elsewhere that depend upon prevention, inhibition, and other forms of negative causation. Our reflection on the importance of passive structures that enable the action of mechanisms suggests that there may be more than one way to lie causally between the inputs and outputs of a mechanism.

Finally, and perhaps most importantly, once we recognize that MIE is but one way to establish constitutive relevance, we open up much needed avenues of research to explore other ways to discover and confirm facts about constitutive relevance. (For useful inroads, see Darden 2006; Darden and Craver 2013; Bechtel & Richardson 1993; Illari 2011; Aizawa and Headley forthcoming).

The core mechanistic ambition to situate phenomena within the causal nexus, showing them to be produced by antecedent causes and underwritten by constitutive mechanisms, continues to animate all sciences, from the cellular physiology of *C. elegans* to the study of black holes. These ambitions are driven in part by the conviction that one who knows these causes and mechanisms knows the buttons and levers one must know to bring the phenomena under our control. The kind of knowledge that allows us to make worms dance, it is hoped, will be the kind of knowledge that will cure schizophrenia or prevent psychiatric disorders. By understanding the logic of this kind of science and its relationship to the causal structure of the world, philosophy can assist in the effort to build a coherent and mutually reinforcing picture both of what the world is like and of how it can most efficiently and decisively be discovered.

## References

- Aizawa, K., & Headley, D. (Forthcoming). Abduction and composition. *Philosophy of Science*.
- Aizawa, K., & Gillett, C. (2016). (Eds.) *Scientific composition and metaphysical ground*. Palgrave-Macmillan.
- Aizawa, K. & Gillett, C. (2019). Defending pluralism about compositional explanations. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 78(101202). <https://doi.org/10.1016/j.shpsc.2019.101202>
- Anscombe, G.E.M. (1971). *Causality and determination: An inaugural lecture*, Cambridge: Cambridge University Press; reprinted in *Metaphysics and the philosophy of mind (The collected philosophical papers of G. E. M. Anscombe, volume 2)* (pp. 133–47). University of Minnesota Press, 1981.
- Baumgartner, M. & Gebharder, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *British Journal for the Philosophy of Science*, 67(3), 731-756.
- Baumgartner, M. & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science*, 84(2), 214-233.
- Bechtel, W. & Richardson, R. C. (2010 [1993]). *Discovering complexity: Decomposition and localization as strategies in scientific research (2<sup>nd</sup> Edition)*. MIT Press/Bradford Books.
- Bechtel, W. (1988). *Philosophy of science: An overview for cognitive science*, Hillsdale, NJ: Erlbaum.

- Bogen, J. (2008). Causally productive activities. *Studies in History and Philosophy of Science Part A*, 39(1), 112-123.
- Couch, M. B. (2011). Mechanisms and constitutive relevance. *Synthese*, 183(3), 375-388.  
doi:10.1007/s11229-011-9882-z.
- Craver, C. F. & Darden, L. (2001). Discovering mechanisms in neurobiology: The case of spatial memory. In P. K. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in neuroscience* (pp. 112-137). University of Pittsburgh Press.
- Craver, C. F. & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.
- Craver, C. F. & W. Bechtel (2007). Top-down causation without top-down causes. *Biology & philosophy*, 22(4), 547–563.
- Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of science*, 68(1), 53-74.
- Craver, C. F. (2005). Beyond reduction: mechanisms, multifield integration and the unity of neuroscience. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 373-395.
- Craver, C. F. (2006). When mechanistic models explain. *Synthese*, 153(3), 355-376.
- Craver, C. F. (2007a). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.

- Craver, C. F. (2007b). Constitutive explanatory relevance. *Journal of Philosophical Research*, 32, 3-20.
- Craver, C. F. (2015). Levels. In T. Metzinger & J. M. Windt (Eds). *Open MIND: 8(T)*. MIND Group.  
doi: 10.15502/978395857049
- Darden, L. (2006). *Reasoning in biological discoveries: Essays on mechanisms, interfield Relations, and anomaly resolution*. Cambridge University Press.
- Dowe, P. (2000). *Physical causation*. Cambridge University Press.
- Dupré, J. (2013). Living Causes. *Aristotelian Society Supplementary Volume 87(1)*, 19–37.
- Dupré, J., & Nicholson, D. J. (2018). A manifesto for a processual philosophy of biology. In D. J. Nicholson & J. Dupré (Eds.), *Everything flows: Towards a processual philosophy of biology* (pp. 1–58). Oxford University Press.
- Ehring, D. (1997). *Causation and persistence*. Oxford University Press.
- Gillett, C. (2016). *Reduction and emergence in science and philosophy*. Cambridge University Press.
- Gebharter, A. (2017). Uncovering constitutive relevance relations in mechanisms. *Philosophical studies* 174(11), 2645-2666.
- Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44(1), 49-71.
- Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of science*, 69(S3), S342-S353.

- Glennan, S. S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntnis*, 72(2), 251-266.
- Glennan, S. S. (2011). Singular and general causal relations: A mechanist perspective. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the Sciences* (pp. 789–817). Oxford University Press.
- Glennan, S. S. (2017). *The new mechanical philosophy*. Oxford University Press.
- Glennan, S. S. (2020). Corporeal composition. *Synthese*. <https://doi.org/10.1007/s11229-020-02805-x>.
- Harbecke, J. (2010). Mechanistic constitution in neurobiological explanations. *International Studies in the Philosophy of Science*, 24(3), 267-285.
- Harinen, T. (2018). Mutual manipulability and causal inbetweenness. *Synthese*, 195(1), 35-54.
- Haugeland, J. (1998). *Having thought*. Harvard University Press.
- Heil, John. (2003). *From an ontological point of view*. Oxford University Press.
- Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, 2(1), 119-135.
- Illari, P. M. (2011). Mechanistic evidence: disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science*, 25(2), 139-157.
- Kaiser, M. I., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *The British Journal for the Philosophy of Science*, 68(3), 745-779.

- Kaiser, M. (2018). The components and boundaries of mechanisms. In S. Glennan & P. M. Illari (Eds.), *The routledge handbook of mechanisms and mechanical philosophy* (pp. 116-130). Routledge.
- Kim, J. (2000). Making sense of downward causation. In P. B. Anderson, C. Emmeche, N. O. Finnemann & P.V. Christiansen (Eds.), *Downward causation* (pp. 305–321). Aarhus University Press.
- Krickel, B. (2018). *The mechanical world: The metaphysical commitments of the new mechanistic approach*. Springer Verlag.
- Ladyman, J., & Ross, D. (2007). *Every thing must go: Metaphysics naturalized*. Oxford University Press on Demand.
- LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and molecular neurobiology*, 23(4), 727-738.
- Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *The British Journal for the Philosophy of Science*, 63(2), 399-427.
- Lewis, D. (2000). Causation as influence. *The Journal of Philosophy*, 97(4), 182-197.
- Lewis, D. (1986). Causal explanation. In D. Lewis (Ed.), *Philosophical papers (volume 2)* (pp. 214-240). Oxford University Press.
- Machamer, P., Darden, L., & Craver, C. F. [MDC] (2000). Thinking about mechanisms. *Philosophy of science*, 67(1), 1-25.

- Mackie, J. L., & Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon Press.
- Nicholson, D. J. (2012). The concept of mechanism in biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1), 152-163.
- Pearl, J. (2012). The mediation Formula: A guide to the assessment of causal pathways in nonlinear models. In C., Berzuini, P., Dawid, & L. Bernardinell (Eds.), *Causality: Statistical perspectives and applications* (pp. 151-179). John Wiley & Sons.
- Prychitko, E. (2019). The causal situationist account of constitutive relevance. *Synthese*, 1-15. <https://dx.doi.org/10.1007/s11229-019-02170-4>
- Reichenbach, H. (1956). *The direction of time*. University of California Press.
- Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese*, 192(11), 3731-3755.
- Salmon, W. (1980). Probabilistic causality. *Pacific Philosophical Quarterly*, 61(1-2), 50-74.
- Salmon, W. C. (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science*, 64(3), 461-477. doi:10.1086/392561
- Salmon, W. (1998). *Causality and explanation*. Oxford University Press. doi:10.1093/0195108647.001.0001
- Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.

- Sanford, D. H. (1993). The problem of the many, many composition questions, and naive mereology. *Noûs*, 27(2), 219-228.
- Scheines, R. (2005). The similarity of causal inference in experimental and non-experimental studies. *Philosophy of Science*, 72(5), 927-940.
- Sengupta, P., & Samuel, A. D. (2009). *Caenorhabditis elegans*: a model system for systems neuroscience. *Current opinion in neurobiology*, 19(6), 637-643.
- Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). *Causation, prediction, and search*. MIT press.
- Stevens, C. F. (1998). A million dollar question: does LTP = memory? *Neuron*, 20(1), 1-2.
- Tabery, J. G. (2004). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of science*, 71(1), 1-15.
- van Eck, D. (2019). Constitutive relevance in cognitive science: The case of eye movements and cognitive mechanisms. *Studies in History and Philosophy of Science Part A*, 73, 44-53.
- Wasserman, R. (2018). Material constitution. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall Edition)*, URL = <https://plato.stanford.edu/archives/fall2018/entries/material-constitution/>.
- Wilhelm, I. (2019). The Ontology of Mechanisms. *The Journal of Philosophy*, 116(11), 615-636.
- Wimsatt, W. C. (1976). Reductionism, levels of organization, and the mind–body problem. In G. Globus, I. Savodnik, & G. Maxwell (Eds.), *Consciousness and the brain* (pp. 199–267). Plenum Press.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford university press.

Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91(2), 303-347.

Woodward, J. (Forthcoming). Causal complexity, conditional independence, and downward causation. *Philosophy of science*.