Constitutive relevance, mutual manipulability, and fat-handedness

Michael Baumgartner · Alexander Gebharter

Abstract

The first part of this paper argues that if Craver's ([2007a], [2007b]) popular mutual manipulability account (MM) of mechanistic constitution is embedded within Woodward's ([2003]) interventionist theory of causation—for which it is explicitly designed—it either undermines the mechanistic research paradigm by entailing that there do not exist relationships of constitutive relevance or it gives rise to the unwanted consequence that constitution is a form of causation. The second part shows how Woodward's theory can be adapted in such a way that (MM) neither undermines the mechanistic paradigm nor reduces constitution to causation. However, it turns out that this modified theoretical embedding of (MM) makes it impossible to produce empirical evidence for constitutive relations. The paper ends by suggesting an additional criterion, the fat-handedness criterion, which, when combined with (MM), generates indirect empirical evidence for constitutive relevance.

Contents

1 Introduction

To mechanistically explain the behavior of a system S, it must be determined which of its parts are constitutively relevant to S's behavior, and what causal structure regulates the behavior of those parts (Craver [2007b], §4). While there are various well-tried frameworks for uncovering causal structures (Spirtes, Glymour, and Scheines [2000]; Pearl [2009], for example), the problem of how to account for relations of constitutive relevance has, at least until Craver's influential book *Explaining the Brain* ([2007b]),

attracted almost no attention. Even though Craver's mutual manipulability approach (MM) to constitutive relevance has been critically examined repeatedly (see, for example, Harbecke [2010]; Couch [2011]; Leuridan [2012]; Glauer [2012], §3.3.2; Schindler [forthcoming]; Franklin-Hall [unpublished]), (MM) is the dominant account of constitutive relevance—notably because of its considerable intuitive plausibility and alleged closeness to scientific practice.[1]

In a nutshell, according to (MM), the behavior of a spatiotemporal part X of a system S is constitutively relevant to S's behavior if, and only if, the behaviors of X and S are mutually manipulable (Craver [2007a], p. 15, p. 17). By 'manipulability' Craver ([2007b], §4.8.3) means the possibility of a change in behavior brought about by an intervention à la Woodward ([2003]). The aim of this paper is to scrutinize to what extent (MM) adequately identifies relations of constitutive relevance.

In the first part, we demonstrate that when (MM) is embedded within Woodward's ([2003]) interventionist theory of causation—for which it is explicitly designed—it either undermines the mechanistic research paradigm by entailing that there do not exist cases of constitution, or it gives rise to the consequence that constitution is a form of causation, a consequence which most mechanists reject. These unwanted ramifications of (MM) are symptoms of a more general problem that Woodward's interventionist theory faces when it is applied to structures containing non-causally dependent variables, such as variables that are related in terms of logical entailment, definition, supervenience, and the like (see, for example, Baumgartner [2010], [2013]).

In a recent paper, Woodward ([2011]) recognizes this problem and proposes a modification of his original theory, which he claims produces adequate results even if applied to structures featuring non-causally dependent variables. In the second part of the paper, we reconsider the prospects of (MM) when embedded in this recent version of interventionism. It turns out that the latter does not solve the problems of (MM). We then propose a further (temporal) refinement of the interventionist theory and show that, when embedded in this temporal variant of interventionism, (MM) neither undermines the mechanistic paradigm nor reduces constitution to causation.

Yet, although (MM) works as intended within this refined interventionist framework, the final part of the paper reveals that the mere mutual manipulability of the behavior of a part X and of the behavior of a macro system

[1] Authors that draw on (MM) include Illari and Williamson ([2011]), Casini *et al.* ([2011]), Soom ([2011]), Kaplan ([2012]) (who is particularly enthusiastic about MM), Gillett ([2013]), Irvine ([2013], §6), or Zednik ([forthcoming]). According to Levy ([2009], 141), (MM) is one of the main achievements of Craver's book. Even Fagan ([2013]), who agues that (MM) has considerable weaknesses, acknowledges that 'Craver's is the best-developed account of MEx [mechanistic explanation] in biology currently on offer' (p. 100).

S does not provide empirical evidence for the former being a constituent of the latter—a finding which contradicts the standard opinion in the literature (see, for example, Craver [2007b], p. 132, p. 159; Couch [2011], p. 381). The paper ends by developing an additional criterion, the fat-handedness criterion. We show that constitution provides the best available explanation for systems satisfying both mutual manipulability and fat-handedness. Mutual manipulability and fat-handedness together, thus, lend abductive support (see, for example, Schurz [2008]) for the existence of constitutional relationships.

2 Mechanisms and constitutive relevance

In recent years, mechanisms have begun to play an increasingly important role in the philosophy of science, especially when it comes to explaining the behavior of macro systems in the special sciences. Among the most frequently cited mechanistic theories are (Machamer, Darden, and Craver [2000]), (Glennan [2002]), and (Bechtel and Abrahamsen [2005]). Craver ([2007a], p. 6) identifies the following consensus among mechanists: 'mechanisms are entities and activities organized such that they exhibit the phenomenon to be explained. The entities are the parts. The activities are what they do.' Plainly, this characterization raises numerous follow-up questions, for instance, as to the nature of the entities and activities that figure in mechanisms. For our subsequent purposes, however, the details of a comprehensive account of the notion of a mechanism are not relevant. In what follows, we sidestep all complications that come with spelling out that notion. Instead, we focus on a simple exemplary mechanism α, whose relevant features are so generic and unspecific that none of our results will hinge on the concrete structure of α. α is depicted in figure 1.

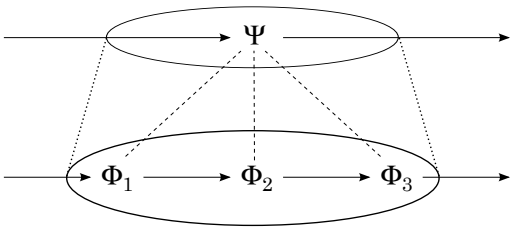

Figure 1: Exemplary mechanism α. Ψ represents the mechanism's behavior at the macro level; Φ_1, Φ_2, and Φ_3 stand for the behaviors of α's constitutively relevant parts. The arrows symbolize direct causation (see §3), the dotted lines indicate that the macro level supervenes on the micro level, and the dashed lines represent constitution. For instance, α can be taken to represent the mechanism constituting the nastic movement of a *Mimosa* (modeled by Ψ), where Φ_1 models the release of potassium ions in the pulvini cells, Φ_2 the cells' turgor pressure, and Φ_3 whether their parenchyma tissue collapses.

α is a whole consisting of several parts. Mechanisms are systems whose behavior can be described at two different levels of specification: the level of the system as a whole (the macro level) and the level of the system's causally interacting constituents (the micro level). The label 'mechanism' is applied both to the whole and to the causal organization of its parts (Craver [2007b], pp. 6–7). A mechanism's macro behavior is often also called a 'phenomenon'. To avoid terminological complications, we subsequently simply speak of a mechanism's macro and micro levels. The macro level of α is represented by the upper, the micro level by the lower ellipse in figure 1.

Macro and micro levels of mechanisms are typically represented by means of individual constants S, X_1, X_2, X_3, ..., where S refers to the system as a whole and X_i refers to a constitutively relevant part of S, and variables Ψ, Φ_1, Φ_2, Φ_3, ..., where the values of Ψ represent the possible behaviors of S and the values of Φ_i the possible behaviors of X_i (Craver [2007b], pp. 153-60). A theory of constitutive relevance must spell out under which conditions X_i's Φ_i-ing, i.e. X_i's realizing of one of the values of Φ_i, is constitutively relevant to S's Ψ-ing, i.e. S's realizing of one of the values of Ψ.

In what follows, we represent 'S's Ψ-ing' by means of the specific variable $\Psi(S)$, whose values from the space $val(\Psi(S))$ correspond to the possible behaviors of the system S as a whole; and 'X_i's Φ_i-ing' we represent by means of the specific variable $\Phi_i(X_i)$, whose values from the space $val(\Phi_i(X_i))$ correspond to the possible behaviors of S's part X_i.[2] Contrary to generic variables that are used to represent behaviors as exhibited by any entities, specific variables represent behaviors as exhibited by particular entities (Spohn [2006]). That is, $\Phi_i(X_i)$ taking one of its values, for instance $\Phi_i(X_i) = \phi_1$, means that entity X_i exhibits behavior ϕ_1, and analogously for $\Psi(S) = \psi_1$. For ease of reading, we will mostly abstain from making the specificity of $\Psi(S)$ and $\Phi_i(X_i)$ explicit; that is, we will simply write Ψ for $\Psi(S)$ and Φ_i for $\Phi_i(X_i)$.

Even though concrete interpretations of the variables in our exemplary mechanism α are not critical for our argument, the latter will be more easily accessible relative to a concrete background. Thus, for instance, $\Psi(S)$ could be taken to represent the nastic movement of a *Mimosa*. The spatiotemporal parts of the *Mimosa* (S) that are involved in this phenomenon are the potassium ions in the pulvini cells (X_1), the latter's turgor pressure (X_2), and the cells' parenchyma tissue (X_3). More specifically, the phenomenon $\Psi(S)$ is constituted by a sudden release of potassium ions in the pulvini cells $\Phi_1(X_1)$, which causes a drop in the cells' turgor pressure $\Phi_2(X_2)$, which, in turn, leads to a collapse of their parenchyma tissue $\Phi_3(X_3)$.[3]

Before we can turn to Craver's mutual manipulability theory of constitution, we have to assemble the essential features of the relation of constitutive relevance. It has three uncontroversial features and one feature that is con-

[2] In principle, the value spaces $val(\Psi(S))$ and $val(\Phi_i(X_i))$ can be infinite.

[3] For a more detailed description see, for instance, (Jaffe, Leopold, and Staples [2002]).

tested. Let us begin with the former. It is generally agreed that the macro behaviors of a mechanism supervene on their constitutively relevant micro behaviors, which amounts to the claim that there cannot be a difference in macro behaviors without a difference in micro behaviors (Craver [2007b], p. 153), or contrapositively put:

(Sup) Every change in the supervening behavior is necessarily accompanied by a change in the supervenience base.

Note that (Sup) only renders it impossible to change the macro level of a mechanism without changing its micro level. By contrast, the latter may well be changeable without a corresponding change in the former.

While the mechanistic literature is clear with respect to (Sup), it is not determinate about what the supervenience base of a mechanism's macro level is. In consequence, it is somewhat dubious what (Sup) entails when applied to mechanism α. Craver ([2007a], p. 15), for example, claims that the macro level of a mechanism 'supervenes on the organized activities of all of the components in the mechanism' (similarly Soom [2012], p. 661). Since the 'activities' in mechanisms are typically represented by causal arrows, this quote suggests that α's macro behavior Ψ supervenes on

(a) the variables in $\mathbf{V}_1 = \{\Phi_1, \Phi_2, \Phi_3\}$ and the causal structure $\Phi_1 \longrightarrow \Phi_2 \longrightarrow \Phi_3$.

In that case, (Sup) entails that any change in Ψ must either be accompanied by a change in the values of at least one variable in $\mathbf{V}_1$ or by a change in the underlying causal structure. Elsewhere, by contrast, Craver ([2007b], p. 153) says that 'supervenience [...] is a relation between a phenomenon and the corporate behavior of the organized components'. As behaviors of components are represented by value configurations of the variables in $\mathbf{V}_1$, this quote suggests that α's macro behavior Ψ supervenes on

(b) the variables in $\mathbf{V}_1 = \{\Phi_1, \Phi_2, \Phi_3\}$.

In that case, it follows from (Sup) that any change in Ψ must be accompanied by a change in at least one of the variables in $\mathbf{V}_1$.

Obviously, candidate supervenience bases (a) and (b) are logically dependent—even more so against the background of a difference-making theory of causation as Woodward's ([2003]), which Craver ([2007b], §3) explicitly endorses. According to Woodward, the causal structuring among a set of variables is determined by the value configurations of those variables under possible interventions (see §3). There cannot be a change in a causal structure without a change in involved variables. Hence, macro phenomena that supervene both on their constituents' behaviors and activities in the vein of (a), supervene on the behavior of their constituents alone—as expressed in (b). That is, by following Craver ([2007b]) in adopting Woodward's interventionist theory of causation as our conceptual background we

can safely conclude that the macro behavior of mechanism α represented by Ψ, in any case, supervenes on the variables in $\mathbf{V}_1$. We shall hence settle for candidate supervenience base (b).

The mechanistic literature is also undecided in regard to the ontological question whether a mechanism's macro properties reductively or non-reductively supervene on their constituents. Only a minority of authors take a clear stance on this issue. For instance, Eronen ([2011], §10, [2012]), Glennan ([1996], pp. 61-2), and Kistler ([2009]) favor a pluralistic ontology that comprises both macro and micro properties, while Fazekas and Kertesz ([2011]) and Soom ([2012]) advance the ontological reducibility of the former to the latter. By contrast, most mechanists sidestep metaphysical issues because they take the framework to be metaphysically neutral (see, for example, Craver [2007b], p. 196).

As the main focus of this paper is on the question of identifying constitutive relevance, we will leave the ontological relation between our exemplary mechanism's macro and micro levels open. Rather, we shall adopt a perspective of causal and mechanistic modeling that views the variables Ψ, Φ_1, Φ_2, and Φ_3 as pure modeling devices. Ψ, Φ_1, Φ_2, and Φ_3 are different random variables such that changes in the values of Ψ supervene on changes in $\mathbf{V}_1 = \{\Phi_1, \Phi_2, \Phi_3\}$.

Apart from (Sup), the relation of constitutive relevance is commonly characterized by spatiotemporal congruence (SC) (Craver [2007a], pp. 5–6) and non-redundancy (NR) (Craver [2007b], p. 157):

(SC) The constituents of a mechanism are spatiotemporal parts of the mechanism's macro level.

(NR) A spatiotemporal part of a mechanism's macro level that under no circumstances makes a difference to the mechanism's macro behavior is not a constituent of that mechanism.

Both of these features would need to be spelled out in more detail. But as neither of them will be at the center of our discussion, we can confine ourselves to referring to further reading: Leuridan ([2012], §6) and Franklin-Hall ([unpublished], §4), for example, expound problems that come with (SC), and Harbecke ([2010], p. 275) and Couch ([2011], §5) provide a specification of (NR) in terms of minimal sufficiency and necessity.

(Sup), (SC), and (NR) are the uncontroversial features of the relation of constitutive relevance. There is one additional feature in regard to which opinions diverge: on the one hand, Craver and Bechtel ([2007]) contend that constitutive relevance is a non-causal form of dependence, on the other, Leuridan ([2012]) argues that it should be understood as a special form of causal dependence. Craver and Bechtel ([2007]) offer two reasons to support its non-causal nature: (i) constitutive relevance holds among wholes and their parts, whereas causal relevance relates mereologically independent

entities; (ii) changes in the macro and micro behaviors of a mechanism temporally overlap, whereas causes temporally precede their effects. The vast majority of mechanists join Craver and Bechtel in taking (i) and (ii) to establish the non-causal nature of constitutive relevance. However, Leuridan ([2012]) succeeds in making a powerful case to the contrary. He argues that if causation is understood along the lines of Woodward's interventionist theory of causation, which is the theory many mechanists adopt (Craver [2007b] in particular), it can be inferred that constitutive relevance is merely a form of bidirectional causal dependence.

Since the question whether constitutive relevance is assumed to differ from causal relevance will be crucial to our argument, we provide that assumption with a label as well:

(NC) Constitutive relevance is a non-causal form of dependence.

As (NC) corresponds to the majority opinion, we will standardly assume (NC). We shall, however, explicitly consider the ramifications of rejecting that assumption.

Overall, thus, we take the relation of constitutive relevance to be characterized by (Sup), (SC), (NR), and (NC). Applied to our exemplary mechanism α in figure 1 that amounts to the following:

(i) Ψ supervenes on $\mathbf{V}_1 = \{\Phi_1, \Phi_2, \Phi_3\}$;
(ii) the variables in $\mathbf{V}_1$ model the behaviors of entities that are spatiotemporal parts of the entity and behaviors modeled by Ψ;
(iii) the variables in $\mathbf{V}_1$ model non-redundant difference-makers of Ψ;
(iv) Ψ and the variables in $\mathbf{V}_1$ are not causally related.

3 Mutual manipulability and interventionism

In light of the previous section, the problem of identifying the constituents of a mechanism can be more specifically characterized as the problem of detecting its spatiotemporal parts that are non-redundant difference-makers of the mechanism's macro behavior such that the latter supervenes on the former. The currently most widely adopted approach for identifying relations of constitutive relevance is Craver's ([2007a], [2007b]) mutual manipulability approach (MM):

> My claim is that to establish that X's Φ-ing is relevant to S's Ψ-ing it is sufficient that one be able to manipulate S's Ψ-ing by intervening to change X's Φ-ing (by stimulating or inhibiting) and that one be able to manipulate X's Φ-ing by manipulating S's Ψ-ing. To establish that a component is irrelevant, it is sufficient to show that one cannot manipulate S's Ψ-ing by intervening to change X's Φ-ing and that one cannot manipulate

> X's Φ-ing by manipulating S's Ψ-ing. (Craver [2007a], p. 17; see also Craver [2007b], p. 159)

(MM) stipulates that if behavior Ψ of S (modeled by $\Psi(S)$) and behavior Φ_i of its spatiotemporal part X_i ($\Phi_i(X_i)$) are mutually manipulable, then $\Phi_i(X_i)$ is constitutively relevant to $\Psi(S)$; and if $\Psi(S)$ and $\Phi_i(X_i)$ are not mutually manipulable, then $\Phi_i(X_i)$ is not constitutively relevant to $\Psi(S)$. Taken in combination, (MM) provides a sufficient and necessary condition for constitutive relevance.[4] Furthermore, Craver ([2007b], §4.8.3) understands manipulability in terms of the existence of a possible ideal intervention as defined by Woodward ([2003], p. 98).[5] In sum, (MM) amounts to this:

(MM) $\Phi_i(X_i)$ is constitutively relevant to $\Psi(S)$ if, and only if, (i) X_i and its behavior Φ_i are spatiotemporal parts of S and its behavior Ψ; (ii) there exists a possible intervention $\mathcal{I}_{\Phi_i} = i_{\Phi_i}$ on $\Phi_i(X_i)$ w.r.t. $\Psi(S)$ that changes $\Psi(S)$'s value (or probability distribution); (iii) there exists a possible intervention $\mathcal{I}_{\Psi} = i_{\Psi}$ on $\Psi(S)$ w.r.t. $\Phi_i(X_i)$ that changes $\Phi_i(X_i)$'s value (or probability distribution).

The basic idea behind (MM) is that the mutual manipulability of S's Ψ-ing and the spatiotemporally overlapping Φ_i-ing of X_i guarantees that X_i's Φ_i-ing is a non-redundant difference-maker of S's Ψ-ing. Moreover, the fact that this difference-making relation holds among spatiotemporally overlapping entities indicates that it is of non-causal nature. Overall, the dependencies that are selected by (MM) are exactly those dependencies that have the features of constitutive relevance pinpointed in the previous section.

(MM) entails that in order to show that Φ_i is a constituent of Ψ, it must be shown that there exists a possible intervention $\mathcal{I}_{\Phi_i} = i_{\Phi_i}$ on Φ_i w.r.t. Ψ that is associated with changes in Ψ and a possible intervention $\mathcal{I}_{\Psi} = i_{\Psi}$ on Ψ w.r.t. Φ_i that is associated with changes in Φ_i. For instance, the release of potassium ions in a *Mimosa's* pulvini cells is exhibited to be constitutively relevant to the *Mimosa's* nastic movement if it is demonstrated that it is possible to intervene both on the pulvini cells and thereby affecting the *Mimosa's* movement and on the *Mimosa* as whole and thereby affecting its pulvini cells.

The most straightforward way of establishing such a possibility is to come up with actual bottom-up and top-down interventions $\mathcal{I}_{\Phi_i} = i_{\Phi_i}$ and $\mathcal{I}_{\Psi} = i_{\Psi}$

[4] Couch ([2011], p. 382) claims that Craver's (MM) only provides a sufficient condition for constitutive relevance. In our view, textual evidence clearly contradicts that assessment.

[5] Craver ([2007b], p. 154) adapts Woodward's definition of an ideal intervention for the context of his presentation. However, as Leuridan ([2012], appendix) has shown that all Craver-interventions are Woodward-interventions, this adaptation does not make a difference to our discussion. We will hence focus on the wording of Woodward's much more well-known definition. For a detailed comparison of Craver's and Woodward's definitions see (Leuridan [2012], §5).

as required by (MM), for all actual interventions are possible interventions. In the case of the *Mimosa*, the introduction of potassium-channel blockers is a candidate bottom-up intervention, whereas touching the *Mimosa* by hand is a candidate top-down intervention. By contrast, subject to (MM), establishing that Φ_i is not a constituent of Ψ is more intricate. To this end, it must be shown that interventions of types $\mathcal{I}_{\Phi_i} = i_{\Phi_i}$ and $\mathcal{I}_{\Psi} = i_{\Psi}$ do not possibly exist. Such a negative existential can only be substantiated inductively or via some a priori (conceptual or theoretical) argument that rules out required interventions.

As indicated before, Woodward's ([2003]) interventionist account of causation, interventionism for short, constitutes the conceptual background of (MM). Interventionism is currently one of the most popular theories of causation. Nonetheless, before we can apply (MM) to our exemplary mechanism α we have to review the main tenets of interventionism, because definitional details will be crucial for our discussion. We suspect that one of the main reasons why so many authors believe that (MM) can be fruitfully put to use when it comes to accounting for constitution is that they rely on a merely intuitive understanding of interventionism without taking the definitional foundation of that theory at face value. For that reason, we will review the relevant definitions and extract their pertinent consequences—even though these definitions are frequently cited in the literature.

The definitions that constitute the core of Woodward's interventionism are the following. (We slightly adapt Woodward's symbolism to avoid confusions with the symbolism used in this paper.)

(M) A necessary and sufficient condition for $\mathcal{X}$ to be a (type-level) direct cause of $\mathcal{Y}$ with respect to a variable set $\mathbf{V}$ is that there be a possible intervention on $\mathcal{X}$ that will change $\mathcal{Y}$ or the probability distribution of $\mathcal{Y}$ when one holds fixed at some value all other variables $\mathcal{Z}_i$ in $\mathbf{V}$. A necessary and sufficient condition for $\mathcal{X}$ to be a (type-level) *contributing cause* of $\mathcal{Y}$ with respect to variable set $\mathbf{V}$ is that (i) there be a directed path from $\mathcal{X}$ to $\mathcal{Y}$ such that each link in this path is a direct causal relationship [...], and that (ii) there be some intervention on $\mathcal{X}$ that will change $\mathcal{Y}$ when all other variables in $\mathbf{V}$ that are not on this path are fixed at some value. (Woodward [2003], p. 59)

(IV) $\mathcal{I}$ is an intervention variable for $\mathcal{X}$ with respect to $\mathcal{Y}$ if and only if [...] (i) $\mathcal{I}$ causes $\mathcal{X}$; (ii) $\mathcal{I}$ acts as a switch for all the other variables that cause $\mathcal{X}$ [...]; (iii) any directed path from $\mathcal{I}$ to $\mathcal{Y}$ goes through $\mathcal{X}$ [...]; (iv) $\mathcal{I}$ is (statistically) independent of any variable $\mathcal{Z}$ that causes $\mathcal{Y}$ and that is on a directed path that does not go through $\mathcal{X}$. (Woodward [2003], p. 98)

Relative to the notion of an intervention variable, an intervention on $\mathcal{X}$ w.r.t. $\mathcal{Y}$ then amounts to an intervention variable $\mathcal{I}$ for $\mathcal{X}$ w.r.t. $\mathcal{Y}$ taking

some value i such that $\mathcal{I} = i$ causes $\mathcal{X}$ to take some value x (Woodward [2003], p. 98).

For our purposes, two things need to be made explicit about a theory built on (M) and (IV). First, the notion of causation provided by (M) is relativized to a set of variables $\mathbf{V}$, but the notion of an intervention variable defined by (IV) is not relativized in that manner (Woodward [2008], p. 202).[6] $\mathcal{X}$ is a direct or contributing cause of $\mathcal{Y}$ only relative to some variable set $\mathbf{V}$, while the interventionist nature of a variable $\mathcal{I}$ does not depend on the set of modeled variables. It must also be noted that the notion of causation that appears in (IV) is not the relativized notion defined in (M), i.e. not the ternary relation '$\mathcal{X}$ causes $\mathcal{Y}$ with respect to $\mathbf{V}$'. Rather, (IV) draws on causation *simpliciter* which Woodward ([2008], p. 209) defines via existential generalization of (M): a variable $\mathcal{X}$ is a cause of $\mathcal{Y}$ if, and only if, there exists a set $\mathbf{V}$ with respect to which $\mathcal{X}$ is a direct or a contributing cause of $\mathcal{Y}$ as defined by (M).

Second, (M) and (IV) establish a tight connection between manipulability, difference-making in context, and causality, which Woodward ([2003], p. 61) sums up in the following slogan: no causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference. In particular, the analysans of causation supplied by (M) stipulates that if $\mathcal{X}$ is a cause of $\mathcal{Y}$, then (i) there exists a possible intervention $\mathcal{I} = i$ on $\mathcal{X}$ with respect to $\mathcal{Y}$ and (ii) all other variables in the relevant variable set $\mathbf{V}$ that are not on a path from $\mathcal{X}$ to $\mathcal{Y}$ can be fixed when $\mathcal{X}$ is wiggled through $\mathcal{I} = i$. That is, (M) determines that the manipulability of $\mathcal{X}$ and the fixability of all off-path variables in $\mathbf{V}$ are each necessary for $\mathcal{X}$ to cause $\mathcal{Y}$.[7]

As we shall see below, Woodward ([2011]) has meanwhile modified his original definitions. In order not to confuse causation and interventions as defined by (M) and (IV) with their counterparts as defined in (Woodward [2011]), we shall subsequently speak of (M)-causation and (IV)-interventions. Moreover, we shall assume that our exemplary mechanism α is completely stable and that interventions on α produce invariant results. That means compensatory responses and indirect interferences are assumed to be canceled by additional interventions and/or the experimental setup—which is a standard assumption the mutual manipulability approach requires (Craver [2007b], pp. 156-7).

We are now in a position to apply (MM), (M), and (IV) to α. (MM) determines that the elements of $\mathbf{V}_1 = \{\Phi_1, \Phi_2, \Phi_3\}$ are constituents of Ψ only if for every $\Phi_i \in \mathbf{V}_1$ there exists a possible intervention $\mathcal{I}_{\Phi_i} = i_{\Phi_i}$ on Φ_i w.r.t. Ψ such that $\mathcal{I}_{\Phi_i} = i_{\Phi_i}$ is associated with changes in Ψ, and for every

[6] If (IV) were relativized like (M), interventionism could not distinguish between difference-making relations that stem from causal dependencies and difference-making relations that are due to common causes (for details, see Baumgartner [2013], pp. 12-3).

[7] For further details on these implications of (M) and (IV) see Baumgartner ([2012]).

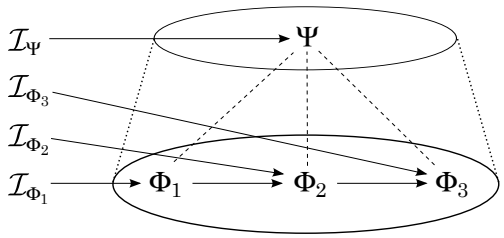

Figure 2: Test design induced by (MM) to identify the constitutive relations between Φ_1, Φ_2, Φ_3 and Ψ.

$\Phi_i \in \mathbf{V}_1$ there exists a possible intervention $\mathcal{I}_\Psi = i_\Psi$ on Ψ w.r.t. Φ_i such that $\mathcal{I}_\Psi = i_\Psi$ is associated with changes in Φ_i. That is, in order for the elements of $\mathbf{V}_1$ to be constituents of Ψ, intervention variables as depicted in figure 2 are required. However, in what follows we prove that such intervention variables are ruled out, and thus, that the elements of $\mathbf{V}_1$ are entailed not to be constituents of Ψ by (MM).

More specifically, we prove that top-down interventions for the macro level of α that are associated with changes in α's micro level cannot possibly exist. To this end, we first identify a condition (†), which, according to (MM), is necessary for the elements of $\mathbf{V}_1 = \{\Phi_1, \Phi_2, \Phi_3\}$ to be constitutively relevant to Ψ and which requires the availability of suitable top-down intervention variables. Second we show that the latter are ruled out and, thus, that condition (†) is unsatisfiable. Here is condition (†):

(†) For every $\Phi_i \in \mathbf{V}_1$ there is an intervention variable $\mathcal{I}_\Psi$ for Ψ w.r.t. Φ_i such that $\mathcal{I}_\Psi = i_\Psi$ (possibly) induces a change in both Ψ and Φ_i.

Now, assume (for *reductio*) that the elements of $\mathbf{V}_1$ are constituents of Ψ. As (†) is necessary for constitution, it follows that for every $\Phi_i \in \mathbf{V}_1$ there is an intervention variable $\mathcal{I}_\Psi$ for Ψ w.r.t. Φ_i. From this it follows by (IV.i) that $\mathcal{I}_\Psi$ is a cause (simpliciter) of Ψ, which, in turn, means that there is a variable set $\mathbf{V}_2$ containing $\mathcal{I}_\Psi$ and Ψ, relative to which $\mathcal{I}_\Psi$ is an (M)-cause of Ψ. According to (M), this entails that there is an intervention variable $\mathcal{I}_{\mathcal{I}_\Psi}$ for $\mathcal{I}_\Psi$ w.r.t. Ψ such that changes induced on $\mathcal{I}_\Psi$ via $\mathcal{I}_{\mathcal{I}_\Psi}$ are associated with changes in Ψ. (SUP) yields that all changes in Ψ are necessarily associated with changes in at least one $\Phi_j \in \mathbf{V}_1$, from which it follows that there is a variable set $\mathbf{V}_3$ containing $\mathcal{I}_\Psi$ and Φ_j w.r.t. which changes induced on $\mathcal{I}_\Psi$ via $\mathcal{I}_{\mathcal{I}_\Psi}$ are associated with changes in Φ_j. Based on this, (M) yields that $\mathcal{I}_\Psi$ is an (M)-cause of Φ_j w.r.t. $\mathbf{V}_3$ and *ipso facto* a cause simpliciter of Φ_j. That is, $\mathcal{I}_\Psi$ is an (M)-cause of both Ψ and Φ_j. Figure 3 depicts the three possible causal structures in which $\mathcal{I}_\Psi$ is a cause of both Ψ and Φ_j.

In (A) and (B) there is a causal relationship between Ψ and Φ_j (which is contained in $\mathbf{V}_1$). Thus, if (NC) is assumed, according to which the elements in $\mathbf{V}_1$ and Ψ are not causally related, it follows that the only structure that

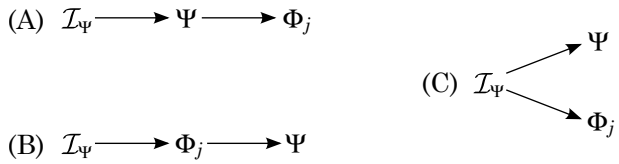

Figure 3: The three possible causal structures in which $\mathcal{I}_\Psi$ is a cause of both Ψ and an element Φ_j of $\mathbf{V}_1$.

can accommodate the fact that $\mathcal{I}_\Psi$ is a cause of both Ψ and Φ_j is (C). That is, (NC) entails that $\mathcal{I}_\Psi$ causes Ψ and Φ_j along two different paths, i.e. that $\mathcal{I}_\Psi$ is a common cause of Ψ and Φ_j.

Now, manipulating Ψ via $\mathcal{I}_\Psi$, which, as shown above, shall be an intervention variable for Ψ w.r.t. Φ_i, either is associated with changes in Φ_i or it is not. If it is, then Φ_i is one of those $\Phi_j \in \mathbf{V}_1$ for which we demonstrated that they are caused by $\mathcal{I}_\Psi$ along a path that differs from the path along which $\mathcal{I}_\Psi$ causes Ψ (see structure (C)). Yet, subject to (IV.iii), intervention variables for Ψ w.r.t. Φ_i must not be common causes of Ψ and Φ_i. That is, satisfying (IV.i) in the vein of (C)—i.e. the only (NC)-compatible way of satisfying (IV.i)—yields that (IV.iii) is violated, which stipulates that all directed causal paths from $\mathcal{I}_\Psi$ to Φ_i must go through Ψ. Thus, $\mathcal{I}_\Psi$ does not comply with all conditions of (IV) and, accordingly, cannot be an intervention variable for Ψ w.r.t. Φ_i after all. By contrast, if manipulating Ψ via $\mathcal{I}_\Psi$ is not associated with changes in Φ_i, then $\mathcal{I}_\Psi$ may count as an intervention variable for Ψ w.r.t. Φ_i but one that is unserviceable to account for a constitutive relation between Ψ and Φ_i. In neither case does $\mathcal{I}_\Psi$ satisfy (†) for Ψ w.r.t. Φ_i.

Nothing in the above reasoning hinges on $\mathcal{I}_\Psi$ being our candidate intervention variable. The same reasoning can be repeated for any other candidate intervention variable. In the structure of mechanism α it is either the case that (IV.i) and (IV.iii) cannot be satisfied together for a triple $\langle \mathcal{I}_\Psi, \Psi, \Phi_i \rangle$, where $\Phi_i \in \mathbf{V}_1$, or, if (IV.i) and (IV.iii) can be jointly satisfied for a triple $\langle \mathcal{I}_\Psi, \Psi, \Phi_i \rangle$, then Φ_i does not change when Ψ is wiggled via $\mathcal{I}_\Psi$. Therefore, (†)—which is necessary for constitution—cannot be satisfied for α. Thus, (MM) entails that the elements of $\mathbf{V}_1$ are not constituents of Ψ—which concludes our *reductio* of the initial assumption that the elements of $\mathbf{V}_1$ are constituents of Ψ.

If we take mechanism α to be instantiated in a *Mimosa's* nastic movement, this *reductio* shows that touching the leaves of the *Mimosa*, for example, does not count as an intervention on the *Mimosa* w.r.t. its pulvini cells, because the touch does not surgically affect the *Mimosa* on the macro level. Rather, it is a common cause of the *Mimosa's* movement and the potassium release in the pulvini cells. The same holds for any other cause of the *Mimosa's* nastic movement. The latter can only be causally affected by

common causes of the *Mimosa's* macro and micro behaviors. None of these common causes are interventions in the sense of (IV). Therefore, there cannot possibly exist any top-down interventions on a *Mimosa's* nastic movement, which, subject to (MM), is thus not constituted by the potassium release in the pulvini cells—nor by any other behavior of its spatiotemporal parts.

Likewise, nothing in the above reasoning hinges on α being our exemplary mechanism. (†) cannot be satisfied for any mechanism. All that is required to establish that there cannot possibly exist an intervention variable as defined by (M) and (IV) for any macro variable Ψ w.r.t. any micro variable Φ_i, such that Φ_i changes when Ψ is wiggled, is that Φ_i is contained in a variable set $\mathbf{V}$, such that Ψ supervenes on $\mathbf{V}$ and that the elements of $\mathbf{V}$ and Ψ are not causally related. That is, as long as constitutive relevance is taken to be characterized by (Sup) and (NC) there cannot possibly exist (M)-(IV)-defined intervention variables for macro variables w.r.t. their micro supervenience bases such that the latter could be changed by intervening on the former.[8] But from the latter finding (MM) infers that no spatiotemporal parts on which a mechanism's macro level supervenes can be constitutively relevant to the latter. In sum, against the background of (Sup), (NC), (M), and (IV), (MM) implies (INEX), *viz.* the inexistence of constitutive relations:

(INEX) There do not exist any relationships of constitutive relevance between macro phenomena and their spatiotemporal parts.

Thus, if (MM) is spelled out in terms of Woodward's original interventionism, for which Craver ([2007b]) explicitly designed it, it reduces the mechanistic paradigm to absurdity.

This *reductio* argument rests on the assumptions (Sup), (NC), (M), and (IV). To avoid its consequence, at least one of these assumptions must be discarded. While (Sup) is uncontested, Leuridan ([2012]) rejects (NC) and Woodward ([2011]) proposes weakened variants of (M) and (IV). We discuss the consequences of rejecting (NC) in the remainder of this section and Woodward's newest variant of interventionism in the next section.

Leuridan's reason for rejecting the non-causal nature of constitution (NC) is that it is a direct consequence of Woodward's interventionist theory that (IV)-manipulability entails causation and, thus, mutual manipulability en-

[8] Based on a problematization of the satisfiability of (IV.iii) that bears some similarity to ours, Glauer ([2012], p. 75) claims, first, that top-down interventions on Ψ w.r.t. Φ_i are not well-defined and, second, that bottom-up interventions on Φ_i w.r.t. Ψ are impossible due to a violation of (IV.iii). We cannot follow Glauer on either count. Relative to a model of a mechanism as in figure 1, it is precisely defined what a top-down intervention amounts to and bottom-up interventions on a mechanism may well exist, for (SUP) only entails that a mechanism's macro level depends on its micro level, but not the other way around. That means it is possible to intervene on the micro level by an (IV)-intervention that is not associated with changes in the macro level. What (SUP) rules out is the existence of (IV)-interventions on the macro w.r.t. to the micro level.

tails bidirectional causation. He argues that Craver cannot consistently embed (MM) within a Woodwardian framework and assume the non-causal nature of constitution as expressed in (NC). Leuridan ([2012], p. 424) concludes: 'Causal relevance (*sensu latu*) thus comes in two varieties: intralevel etiological relevance (or causal relevance *sensu strictu*) and interlevel relevance'.

We agree with Leuridan that, if constitutive relevance is not assumed to be of non-causal nature, (MM), interpreted against the background of (M) and (IV), implies that constitutive relevance reduces to a form of causal relevance. Moreover, discarding (NC) yields that the *reductio* argument leading to (INEX) is blocked, such that cases of constitution—or rather constitutive causation—are no longer entailed to be inexistent on mere conceptual grounds. To see this, note first that if we discard (NC) it still holds in mechanism α that all causes of Ψ are also causes of some $\Phi_i \in \mathbf{V}_1$, for due to (Sup) all changes induced on Ψ are associated with changes in some $\Phi_i \in \mathbf{V}_1$. But by abandoning (NC) it no longer follows that Ψ and $\Phi_i \in \mathbf{V}_1$ must be located on different causal paths. Rather, it becomes possible that Ψ and Φ_i are located on one and the same causal path. If Ψ and some $\Phi_i \in \mathbf{V}_1$ are on the same path, (IV.i) and (IV.iii) are rendered satisfiable at the same time. It thereby becomes possible to intervene on Ψ w.r.t. Φ_i in the sense defined by (IV), such that manipulations of Ψ are associated with changes in Φ_i. Hence, condition (†) turns out to be satisfiable.

If (and only if) for every element Φ_i of $\mathbf{V}_1$ there exist possible interventions of type (A) on Ψ w.r.t. Φ_i (see figure 3), such that Φ_i changes when Ψ is wiggled, and there exist possible interventions of type (B) on Φ_i w.r.t. Ψ, such that Ψ changes when Φ_i is wiggled, (MM) yields that the elements of $\mathbf{V}_1$ and Ψ are related in terms of constitution (or constitutive causation). Relationships of constitutive causation are always bidirectional, i.e. they amount to causal loops. Overall, if Ψ and the elements of $\mathbf{V}_1$ are mutually manipulable, the causal structure entailed by (Sup), (M), (IV), and (MM) for α is the one depicted in figure 4.

The resulting causal structure features numerous causal cycles. Note that the causal dependencies in figure 4 hold among wholes and their parts, that changes in Ψ and changes in $\mathbf{V}_1$ temporally overlap, and that the causal in-

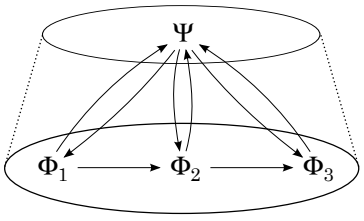

Figure 4: Causal structure underlying mechanism α, provided that constitution is not assumed to be of non-causal nature.

fluence is transmitted instantaneously along the interlevel paths. All of these features of the structure in figure 4 are very unusual for causal structures. If (NC) is discarded, the resulting structures of constitutive causation differ from ordinary causal structures in numerous crucial respects. In light of this, the vast majority of mechanists do not join Leuridan ([2012]) in abandoning (NC). Fortunately, there is another way to block the *reductio* argument leading to (INEX).

4 Modifying interventionism

The problem of (MM) revealed in the previous section is but one problem among many that Woodward's ([2003]) theory gives rise to when applied to variables that are dependent in non-causal ways (for another such problem, see Baumgartner [2010], [2013]). Woodward ([2011]) has recently offered a modified version of interventionism, which he claims avoids the problems associated with non-causal dependence relations. In this section, we reconsider (MM) when read against the background of this most recent proposal.

Woodward's ([2011]) basic idea for modifying his original definitions is to introduce exemption clauses for non-causal dependence relations. He proposes the amended definitions (M*) and (IV*) given below. In both of them, the notion of supervenience appears as proxy for all other non-causal dependence relations. For brevity, we merge Woodward's twofold account of direct and contributing causation into one notion of causation (M*):

(M*) $\mathcal{X}$ is a cause of $\mathcal{Y}$ w.r.t. variable set $\mathbf{V}$ if, and only if, there possibly exists an (IV*)-intervention on $\mathcal{X}$ w.r.t. $\mathcal{Y}$, when all variables in $\mathbf{V}$ that are not on a causal path from $\mathcal{X}$ to $\mathcal{Y}$ and are not related in terms of supervenience to $\mathcal{X}$ or $\mathcal{Y}$ are held fixed, such that the value or the probability distribution of $\mathcal{Y}$ changes.[9]

(IV*) $\mathcal{I}$ is an intervention variable for $\mathcal{X}$ w.r.t. $\mathcal{Y}$ if, and only if, $\mathcal{I}$ satisfies (IV.i), (IV.ii), (IV.iii*), and (IV.iv*):

- (IV.iii*) any directed path from $\mathcal{I}$ to $\mathcal{Y}$ goes through $\mathcal{X}$ or through a variable $\mathcal{Z}$ which is related to $\mathcal{X}$ in terms of supervenience;
- (IV.iv*) $\mathcal{I}$ is (statistically) independent of every cause of $\mathcal{Y}$ which is neither located on a path through $\mathcal{X}$ nor on a path through a variable $\mathcal{Z}$ which is related to $\mathcal{X}$ in terms of supervenience.

To distinguish this modified variant of interventionism from the original variant, we shall speak of (M*)-causation and (IV*)-interventions and we shall say that (M*) and (IV*) constitute interventionism*.

[9] Woodward ([2011]) does not state (M*) explicitly but merely indicates its relevant features. (M*) is our reconstruction of Woodward's suggestion. For a justification as to why this is an adequate reconstruction see Baumgartner ([2013], §3).

The relevant respects (for our purposes) in which the exemption clauses in (M*) and (IV*) return definitions that differ from (M) and (IV) are as follows. First, according to (IV*), a variable $\mathcal{I}$ can count as an intervention variable for $\mathcal{X}$ w.r.t. $\mathcal{Y}$ even if $\mathcal{I}$ is connected to $\mathcal{Y}$ through a path that does not go through $\mathcal{X}$, i.e. through what we shall call an $\overline{\mathcal{X}}$-path, provided that this $\overline{\mathcal{X}}$-path goes through a variable contained in the supervenience base of $\mathcal{X}$. Second, (M*) entails that in order to establish that $\mathcal{X}$ is a cause of $\mathcal{Y}$, those variables that are contained in the supervenience bases of $\mathcal{X}$ and $\mathcal{Y}$ do not have to be fixed.

Let us now see what the implications of (MM) are when read against the background of interventionism*. Again, we first apply (MM) to mechanism α and assume all the features assigned to constitution in section 2, and we once more assume that the variables in $\mathbf{V}_1 = \{\Phi_1, \Phi_2, \Phi_3\}$ are constituents of Ψ. From this, it follows by (MM)—more specifically, by condition (†) entailed by (MM)—that for every $\Phi_i \in \mathbf{V}_1$ there is an intervention variable $\mathcal{I}_\Psi$ for Ψ w.r.t. Φ_i. $\mathcal{I}_\Psi$ is a cause of Ψ, which means there is a variable set $\mathbf{V}_2$ containing $\mathcal{I}_\Psi$ and Ψ, relative to which there is an intervention variable $\mathcal{I}_{\mathcal{I}_\Psi}$ for $\mathcal{I}_\Psi$ w.r.t. Ψ such that changes induced on $\mathcal{I}_\Psi$ via $\mathcal{I}_{\mathcal{I}_\Psi}$ are associated with changes in Ψ. Due to (Sup), all changes in Ψ are necessarily associated with changes in at least one $\Phi_j \in \mathbf{V}_1$, meaning there is a variable set $\mathbf{V}_3$ relative to which $\mathcal{I}_\Psi$ is an (M*)-cause and, thus, a cause (simpliciter) of Φ_j. That is, $\mathcal{I}_\Psi$ is a cause of both Ψ and Φ_j. Of all possible causal structures in figure 3 that feature $\mathcal{I}_\Psi$ as cause of both Ψ and some elements of $\mathbf{V}_1$ only (C) is compatible with (NC). Hence, $\mathcal{I}_\Psi$ must be a common cause of Ψ and Φ_j.

Up to this point the consequences of (MM) when embedded in interventionism* are identical to its consequences when embedded in interventionism. But while $\mathcal{I}_\Psi$ in structure (C) of figure 3 cannot be an (IV)-intervention variable for Ψ w.r.t. those elements of $\mathbf{V}_1$ that change due to (Sup) when intervening on Ψ, $\mathcal{I}_\Psi$ in (C) can be such an (IV*)-intervention variable for Ψ. The reason is that (IV*) allows for $\overline{\Psi}$-paths from $\mathcal{I}_\Psi$ to elements of $\mathbf{V}_1$, provided that these $\overline{\Psi}$-paths go through the supervenience base of Ψ—which is the case in structure (C). In consequence, Ψ is manipulable via an (IV*)-intervention on Ψ w.r.t. some $\Phi_i \in \mathbf{V}_1$ such that Φ_i changes its value. Moreover if it turns out that for every $\Phi_i \in \mathbf{V}_1$ we find an (IV*)-intervention of type (C) on Ψ w.r.t. Φ_i such that Φ_i changes, Ψ is manipulable w.r.t. every element of $\mathbf{V}_1$. Finally, it is also possible that for every $\Phi_i \in \mathbf{V}_1$ we find an (IV*)-intervention on Φ_i w.r.t. Ψ such that Ψ changes. In that case, (MM) is applicable and identifies the elements of $\mathbf{V}_1$ as constituents of Ψ. So far so good.

However, note that if Ψ and the elements of $\mathbf{V}_1$ turn out to be mutually manipulable, not only is (MM) applicable, but also (M*). If Ψ is (IV*)-manipulable w.r.t. every $\Phi_i \in \mathbf{V}_1$ and every $\Phi_i \in \mathbf{V}_1$ is (IV*)-manipulable w.r.t. Ψ, (M*) rules that there is a causal feedback between Ψ and every

$\Phi_i \in \mathbf{V}_1$. The reason is that (M*) does not require supervenience bases to be held fixed when macro variables are manipulated. Yet, (NC) entails that Ψ is causally independent of all its constituents in $\mathbf{V}_1$. In sum, applying (MM) to α against the background of (Sup), (NC), (M*), and (IV*) results in a contradiction: Ψ and the elements of $\mathbf{V}_1$ are causally dependent, which is entailed by (M*), and not causally dependent, which is entailed by (MM) and (NC).

Again, nothing in the above reasoning hinges on our particular choice of variables or on the concrete structure of our exemplary mechanism. (Sup), (M*), and (IV*) yield that macro variables and their constituents are causally dependent, whereas (MM) and (NC) entail that they are independent. Embedding (MM) in interventionism* reduces the mechanistic framework to absurdity in a way that is closely related to the *reductio* of (MM) when embedded in interventionism. While, in the latter case, assuming the variables in $\mathbf{V}_1$ to be constituents of Ψ entails that they are not constituents of Ψ, the same assumption, in the former case, entails that the elements of $\mathbf{V}_1$ are causally relevant to Ψ and not causally relevant to Ψ. As in the latter case, not only the non-causal nature of constitution, but also the assumption of the existence of constitutive relationships is a candidate for rejection when (MM) is embedded in interventionism*. But discarding the existence of constitutive relationships would just yield (INEX) again. If we want to insist that mechanists do not chase a chimera and if we want constitution to be a non-causal form of dependence, modifying (M) and (IV) in terms of (M*) and (IV*) does not solve the problem revealed in the previous section. Interventionist definitions must be modified further.

One may conceive of several supplementary modifications of interventionist definitions. In what follows, we propose our preferred option without claiming that this is the only way to go. As we want to uphold (MM), we have to block the second *reductio* argument presented above by preventing that mutual manipulability of a mechanism's macro and micro levels entails (mutual) causation. A straightforward way to do so is by stipulating that spatiotemporally overlapping entities and their behaviors are not causally related, even if they are mutually manipulable. We propose to build a constraint into (M*) to the effect that, in order for $\mathcal{X}$ to be a cause of $\mathcal{Y}$, it must be possible to intervene on $\mathcal{X}$ to change the value of $\mathcal{X}$ at a time t such that $\mathcal{Y}$ changes at a (strictly) later time t':[10]

(M)** $\mathcal{X}$ is a cause of $\mathcal{Y}$ w.r.t. variable set $\mathbf{V}$ if, and only if, there exist times t, t', where $t < t'$, such that there possibly exists an (IV*)-intervention on $\mathcal{X}$ w.r.t. $\mathcal{Y}$ that sets $\mathcal{X}$ to some value x at t, when all variables in

[10] We do not want to claim that (M**) amounts to an account of causation that produces adequate results in every conceivable context of application. Our claim is merely that if one wants to embed (MM) in a variant of interventionism that accounts for causation as it figures in mechanisms, (M**) is a promising way to go.

> **V** that are not on a causal path from $\mathcal{X}$ to $\mathcal{Y}$ and are not related in terms of supervenience to $\mathcal{X}$ or $\mathcal{Y}$ are held fixed, such that the value or the probability distribution of $\mathcal{Y}$ changes at t'.

t and t' must be understood to stand for more or less extended time intervals such that t is strictly before t'. Note that (M**) is only a sparse modification of (M*): it does not stipulate that changes in causes always strictly precede changes in their effects. Rather, it merely requires that at least one possible intervention on a cause induces a change in that cause which strictly precedes a change in the corresponding effect. Nonetheless, such a sparse modification suffices to guarantee that relationships of mechanistic constitution are entailed to be of non-causal nature.

To see this, reconsider our exemplary mechanism α. As before, variables that are common causes of Ψ and of some $\Phi_i \in \mathbf{V}_1$ count as (IV*)-intervention variables for Ψ w.r.t. Φ_i. That is, even if for all $\Phi_i \in \mathbf{V}_1$ it holds that every intervention variable $\mathcal{I}_\Psi$ on Ψ w.r.t. Φ_i is a common cause of Ψ and Φ_i, (†) is satisfiable for α. Accordingly, Ψ and the elements of $\mathbf{V}_1$ are rendered mutually manipulable, which, according to (MM), induces constitution. At the same time, however, that particular form of mutual manipulability does not satisfy the right-hand side of (M**), for, according to (SC), all changes induced on Ψ temporally overlap with the changes in Φ_1 to Φ_3. (For instance, the drop in turgor pressure of a *Mimosa's* pulvini cells always coincides with the nastic movement of the plant.) Or the other way around: for no $\Phi_i \in \mathbf{V}_1$ there exist two strictly sequential times t and t' such that Ψ is changed at t and Φ_i changes at t' or Φ_i is changed at t and Ψ changes at t'. (M**) thus entails that Ψ is causally independent of all variables in $\mathbf{V}_1$. These findings generalize for all mechanisms and all chosen variable sets: as (SC) is assumed to hold for all constitutive relations, the right-hand side of (M**) is unsatisfiable, in principle, by macro and micro levels of mechanisms.

Overall, if we embed (MM) in interventionism**, we—for the first time—get a consistent account of mechanistic constitution, which does not force (INEX) upon us. What is more, the resulting account entails the non-causal nature of constitution (NC). Cashing out (MM) within interventionism** dispenses us from additionally assuming (NC). Both of these features are clear advantages the aggregate of (MM) and interventionism** has over any other currently available interventionist embedding of Craver's mutual manipulability criterion.

5 Fat-handedness

Despite the progress we have made so far, there remains a severe problem. It is one of the main selling points of (MM) that it allegedly reproduces experimental practices in the sciences (Craver [2007b], §4). (MM) is taken

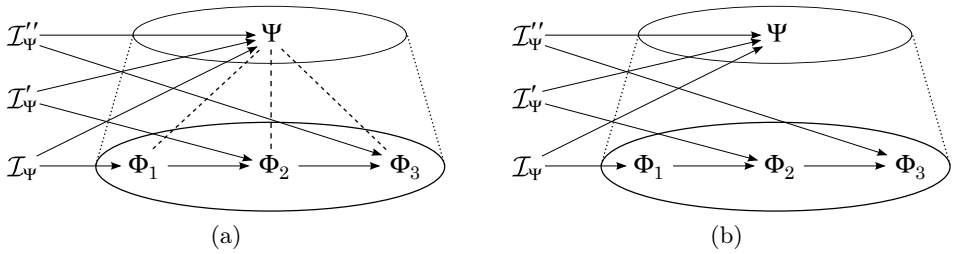

Figure 5: Two empirically indistinguishable models of mutual (IV*)-manipulations of α's macro and micro levels.

to theoretically ground a method for empirically detecting relationships of constitutive relevance. Applied to our exemplary mechanism α that amounts to the claim that the observed mutual manipulability of Ψ and the variables in $\mathbf{V}_1$ provides empirical evidence in favor of the variables in $\mathbf{V}_1$ being the constituents of Ψ.

Yet, while the embedding of (MM) in interventionism**, unlike the original embedding in interventionism, generates a consistent theory, it fails to account for mechanistic constitution on evidence-based grounds. To see this, reconsider mechanism α. In order to yield that the elements of $\mathbf{V}_1$ are constituents of Ψ, (MM) requires that for every $\Phi_i \in \mathbf{V}_1$ there exists an intervention on Ψ w.r.t. Φ_i as well as an intervention on Φ_i w.r.t. Ψ such that these interventions are associated with changes in both Ψ and Φ_i. In virtue of (M**), all of these interventions are causes of both Ψ and Φ_i. As the relation between Ψ and the elements of $\mathbf{V}_1$ is determined to be of non-causal nature by (M**), all of these interventions required to satisfy mutual manipulability are entailed to be common causes of Ψ and Φ_i. (MM) can only be satisfied by fat-handed interventions[11] on macro and micro levels, which only count as interventions because (IV*) has built-in exemption clauses for supervenience relationships. In light of this result, however, the correlations of Ψ and the elements of $\mathbf{V}_1$ that result from mutual manipulability do not need to be due to constitutive dependencies. These correlations could simply be due to the fact that all (MM)-induced interventions on α are fat-handed.

More concretely, figures 5(a) and 5(b) depict the type of (IV*)-interventions that give rise to the mutual manipulability of the macro and micro levels of mechanism α. If we find correlations of Ψ and the micro variables in $\mathbf{V}_1$ under such fat-handed manipulations, there is no need to postulate constitutive dependencies. Such correlations can be accounted for by the mere fact that the macro and micro levels of α are systematically wiggled with a fat hand. Model 5(a), which features constitutive dependencies among Ψ and the elements of $\mathbf{V}_1$, and model 5(b), which does not feature such

[11] A fat-handed intervention is an intervention that influences its ultimate effect along two (or more) different causal paths (Scheines [2005], p. 932).

dependencies, imply the same correlations under manipulations. They are empirically indistinguishable.

This result again generalizes for all mechanisms. (M**) entails that macro and micro levels are only mutually manipulable via common causes. But then, the fact that they behave in a highly correlated manner can be explained by the mere fat-handed nature of corresponding mutual manipulations. Mutual manipulability via common cause interventions provides no empirical evidence in favor of the existence of constitutive dependencies. Thus, (MM) is not sufficient to account for constitution on evidence-based grounds.

In the remainder of this paper, we make a suggestion how (MM) could be supplemented in a way that produces at least indirect empirical evidence—or second-order evidence—in favor of constitutive dependencies. To generate evidence for the existence of an additional dependence among macro and micro levels of a mechanism, top-down interventions are of core relevance. There exists an important asymmetry between bottom-up and top-down interventions: while it is possible to intervene on the micro level of a mechanism without inducing changes on the macro level, (Sup) renders it impossible to change the macro level of a mechanism without thereby inducing changes on the micro level. Subject to (Sup) and (M**), every cause of a mechanism's macro level is a common cause of the macro level and at least one element of the corresponding supervenience base.

In case of mechanism α that means that every cause of Ψ is a common cause of Ψ and at least one $\Phi_i \in \mathbf{V}_1$. Now, suppose that we find a first cause $\mathcal{I}_\Psi$ of Ψ and suppose that $\mathcal{I}_\Psi$ turns out to be a (direct) common cause of Ψ and Φ_1. (See, again, figure 5 for an illustration.) Next, suppose that we expand the analysed variable set and that we find a second cause $\mathcal{I}'_\Psi$ of Ψ which happens to be a (direct) common cause of Ψ and Φ_2. Furthermore, a third cause $\mathcal{I}''_\Psi$ of Ψ turns out to be a (direct) common cause of Ψ and Φ_3, a fourth $\mathcal{I}'''_\Psi$ a (direct) common cause of Ψ and Φ_1, and so on; to the effect that all causes of Ψ are common causes of type $\mathcal{I}_\Psi$, $\mathcal{I}'_\Psi$, or $\mathcal{I}''_\Psi$ in figure 5. The resulting correlations can either be modeled by means of a structure featuring constitutive dependencies, as depicted in figure 5(a), or by means of a pure common cause model, as depicted in figure 5(b).

Even though models 5(a) and 5(b) are equivalent w.r.t. implied correlations, model 5(a) has a distinctive advantage over model 5(b): 5(a) not only explains the correlations of Ψ and the corresponding variables in $\mathbf{V}_1$, it also explains why we do not find causes of Ψ that are not common causes of Ψ and some $\Phi_i \in \mathbf{V}_1$. Model 5(b) only accounts for the correlations of Ψ and the corresponding variables in $\mathbf{V}_1$, but it provides no account of why we do not at some point find a surgical cause of Ψ. If the underlying causal structure were a mere common cause structure expansions of analysed variable sets should be expected to yield causes of Ψ that are not common causes of Ψ and some $\Phi_i \in \mathbf{V}_1$. However, if the elements of $\mathbf{V}_1$ are constituents of Ψ,

there cannot exist a surgical cause of Ψ. The constitutive model 5(a) has higher explanatory power than the common cause model 5(b). It not only explains resulting correlations but it also explains why there do not exist surgical causes of Ψ.

Empirical evidence not only consists in correlational evidence resulting from suitable manipulations. Expansions of analysed variable sets give rise to a sort of second-order evidence. If it turns out that all causes of Ψ we find are common causes of Ψ and some $\Phi_i \in \mathbf{V}_1$, we have good abductive reasons to prefer models that account for this second-order evidence by introducing additional dependencies among Ψ and the variables in $\mathbf{V}_1$. Models featuring constitutive relevance relations provide the best available explanation for an (MM)-satisfying system whose macro behavior can only be manipulated with a fat hand. That is, in order to establish that the elements of $\mathbf{V}_1$ are constitutively relevant to Ψ, we not only have to establish that Ψ and the variables in $\mathbf{V}_1$ are mutually manipulable, but we moreover have to establish that all causes of Ψ are common causes of Ψ and some $\Phi_i \in \mathbf{V}_1$.

Hence, the empirical evidence that justifies to identify the processes in a *Mimosa's* pulvini cells as constituents of the plant's nastic movement does not simply consist in singular mutual manipulations of the *Mimosa's* macro and micro levels. The result of any single such experimental manipulation can be thoroughly accounted for in terms of a pure common cause model. Rather, what empirically justifies the inference to constitutive relevance is the fact that all further macro causes of the *Mimosa's* nastic movement that we happen to find in additional experimental setups turn out to be common causes of the *Mimosa's* macro behavior and some micro occurrence in its pulvini cells, i.e. the fact that no cause of the nastic movement can be found that does not affect the pulvini cells in some way or another.

To put this in more explicit and general terms we introduce the criterion of fat-handedness (FH):

(FH) The elements of a set $\mathbf{V} = \{\Phi_1, \Phi_2, \ldots, \Phi_n\}$ and a variable Ψ satisfy (FH) if, and only if, every (IV*)-intervention on Ψ is a common cause of Ψ and some $\Phi_i \in \mathbf{V}$.

We submit that the conjunction of (MM) and (FH)—embedded in interventionism**—amounts to a consistent account of constitutive relevance that accounts for constitutive relevance on evidence-based grounds:

(CR) The elements of a set $\mathbf{V} = \{\Phi_1, \Phi_2, \ldots, \Phi_n\}$ are constitutively relevant to a variable Ψ if, and only if, the relationship between the elements of $\mathbf{V}$ and Ψ complies with (MM) and (FH).

In sum, to empirically establish relations of constitutive relevance, establishing the mutual manipulability of a variable Ψ and the variables in a set $\mathbf{V}$ on the basis of a few (IV*)-interventions is not sufficient. Additionally,

it must be shown that the dependencies between Ψ and $\mathbf{V}$ induced by interventions on Ψ cannot be screened off by suitable surgical interventions. Only after a systematic expansion of the analysed variable set that generates a sufficient second-order evidential basis to inductively infer that Ψ and the elements of $\mathbf{V}$ satisfy (FH) can it justifiably be concluded that Ψ and $\mathbf{V}$'s elements are related in terms of constitutive relevance. It must be emphasized that the universal logical form of (FH), which imposes constraints on every (IV*)-intervention on Ψ w.r.t. some Φ_i, yields that compliance with (CR) can, in principle, only be established inductively and is, thus, always prone to error.

6 Conclusion

The first part of this paper has shown that Craver's ([2007a], [2007b]) mutual manipulability account of mechanistic constitution requires a theoretical embedding that is much more difficult to come by than is recognized by most mechanists. Embedding (MM) either in Woodward's ([2003]) original variant of interventionism—as suggested by Craver himself—or in Woodward's ([2011]) recent interventionism*, while at the same time assuming the non-causal nature of constitution (NC), undermines the mechanistic framework by implying the inexistence of constitutive relationships (INEX). One way out of the impasse would be to reject (NC), that is, to accept, as does Leuridan ([2012]), that constitution is a special form of causal dependence. However, most friends of mechanistic constitution join Craver and Bechtel ([2007]) in contending that constitution differs in essential respects from causal dependence. That, in turn, means that the interventionist theory in which (MM) is embedded must be modified further.

In the second part, we proposed a temporal relativization of (M*), *viz.* (M**), which, for the first time, resulted in a consistent theoretical embedding of (MM) that neither forces (INEX) nor the rejection of (NC) upon us. Still, it turned out that this embedding in interventionism** falls short of accounting for constitutive relationships on evidence-based grounds, which, after all, is believed to be one of the core advantages of (MM). We hence suggested to use unresolvable fat-handedness of top-down interventions as an additional criterion (FH), which, in combination with (MM), allows for generating (second-order) abductive evidence for constitutive dependencies.

Let us end with a caveat. We do not claim that the combination of (MM), interventionism**, and (FH) accounts for how the notion of constitutive relevance is actually used in scientific practice. Establishing this would require successful replications of case studies based on our account, which is a project we have not even begun to undertake here. Rather, we contend that embedding (MM) in interventionism** and supplementing it with (FH) constitutes a consistent and empirically contentful account of mechanistic

constitution, which allows for treating constitution and causation as two closely related, but still distinct types of dependencies. According to this account, both constitution and causation are difference-making relations that stem from manipulability and control; but while the former is characterized by unresolvable fat-handedness and systematic temporal overlap, the latter allows for surgical interventions and, at least sometimes, features strict sequentiality.

The final upshot of this paper is a call for philosophical caution. Craver's mutual manipulability account of constitution and Woodward's interventionist account of causation have a considerable amount of pre-theoretic intuitive plausibility. They both seem to directly capture scientific practice. But intuitive appeal and echoing scientific practice do not suffice for a fruitful, not even for a consistent, philosophical theory.

Michael Baumgartner:
Department of Philosophy
University of Geneva
2, Rue de Candolle
1211 Geneva
Switzerland
e-mail: michael.baumgartner@unige.ch

Alexander Gebharter:
Düsseldorf Center for Logic and Philosophy of Science
Heinrich Heine University Düsseldorf
Universitätsstraße 1
40225 Düsseldorf
Germany
e-mail: alexander.gebharter@phil.hhu.de

Funding

German Research Foundation (FOR 1063 to A.G.).

Acknowledgements

We thank Markus Eronen, Jens Harbecke, Lena Kästner, Bert Leuridan, Gerhard Schurz, Patrice Soom, and Carlos Zednik for valuable discussions

about the arguments presented in this paper as well as one editor and three anonymous referees for helpful comments on earlier drafts. Moreover, we have benefited from feedback from the LeRM group in Osnabrück and from the audiences of GWP.2013 in Hannover and of EPSA13 in Helsinki. We thank Alexander Christian for help with the *Mimosa* example and Sebastian Maaß and Alexander G. Mirnig for proofreading.

References

Baumgartner, M. [2010]: 'Interventionism and Epiphenomenalism', *Canadian Journal of Philosophy*, **40**, pp. 359–84.

Baumgartner, M. [2012]: 'The Logical Form of Interventionism', *Philosophia*, **40**, pp. 751–61.

Baumgartner, M. [2013]: 'Rendering Interventionism and Non-Reductive Physicalism Compatible', *Dialectica*, **67**, pp. 1–27.

Bechtel, W. and Abrahamsen, A. [2005]: 'Explanation: A Mechanist Alternative', *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, **36**, pp. 421–41.

Casini, L., Illari, P. M., Russo, F. and Williamson, J. [2011]: 'Models for Prediction, Explanation and Control: Recursive Bayesian Networks', *Theoria*, **26**, pp. 5–33.

Couch, M. B. [2011]: 'Mechanisms and Constitutive Relevance', *Synthese*, **183**, pp. 375–88.

Craver, C. [2007a]: 'Constitutive Explanatory Relevance', *Journal for Philosophical Research*, **32**, pp. 3–20.

Craver, C. [2007b]: *Explaining the Brain*, Oxford: Clarendon Press.

Craver, C. and Bechtel, W. [2007]: 'Top-down Causation Without Top-down Causes', *Biology and Philosophy*, **22**, pp. 547–63.

Eronen, M. [2011]: *Reduction in Philosophy of Mind: A Pluralistic Account*, Frankfurt (Main): Ontos.

Eronen, M. [2012]: 'Pluralistic Physicalism and the Causal Exclusion Argument', *European Journal for the Philosophy of Science*, **2**, pp. 219–32.

Fagan, M. B. [2013]: *Philosophy of Stem Cell Biology. Knowledge in Flesh and Blood*, New York: Palgrave Macmillan.

Fazekas, P. and Kertesz, G. [2011]: 'Causation at Different Levels: Tracking the Commitments of Mechanistic Explanations', *Biology and Philosophy*, **26**, pp. 365–83.

Franklin-Hall, L. R. [unpublished]: 'The Emperor's New Mechanisms', <`https://files.nyu.edu/lrf217/public/scientific-explanation/franklin-hall----the-empero.pdf`>

Gillett, C. [2013]: 'Constitution, and Multiple Constitution, in the Sciences: Using the Neuron to Construct a Starting Framework', *Minds and Machines*, **23**, pp. 309–37.

Glauer, R. D. [2012]: *Emergent Mechanisms. Reductive Explanation for Limited Beings*, Münster: Mentis.

Glennan, S. [1996]: 'Mechanisms and the Nature of Causation', *Erkenntnis*, **44**, pp. 49–71.

Glennan, S. [2002]: 'Rethinking Mechanistic Explanation', *Philosophy of Science*, **69**, pp. S342–53.
Harbecke, J. [2010]: 'Mechanistic Constitution in Neurobiological Explanations', *International Studies in the Philosophy of Science*, **24**, pp. 267–85.
Illari, P. M., Russo, F. and Williamson, J. [2011]: *Causality in the Sciences*, Oxford: Oxford University Press.
Illari, P. M. and Williamson, J. [2011]: 'Mechanisms are Real and Local', in P. M. Illari, F. Russo and J. Williamson (eds), *Causality in the Sciences*, Oxford: Oxford University Press, pp. 818–44.
Irvine, E. [2013]: *Consciousness as a Scientific Concept*, Dordrecht: Springer.
Jaffe, J. J., Leopold, A. C. and Staples, R. C. [2002]: 'Thigmo Responses in Plants and Fungi', *American Journal of Botany*, **89**, pp. 375–82.
Kaplan, D. M. [2012]: 'How to Demarcate the Boundaries of Cognition', *Biology and Philosophy*, **27**, pp. 545–70.
Kistler, M. [2009]: 'Mechanisms and Downward Causation', *Philosophical Psychology*, **22**, pp. 595–609.
Leuridan, B. [2012]: 'Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms', *British Journal for the Philosophy of Science*, **63**, pp. 399–427.
Levy, A. [2009]: 'Carl F. Craver, Explaining what? Review of Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience', *Biology and Philosophy*, **24**, pp. 137–45.
Machamer, P., Darden, L. and Craver, C. [2000]: 'Thinking about mechanisms', *Philosophy of Science*, **67**, pp. 1–25.
Pearl, J. [2009]: *Causality*, Cambridge: Cambridge University Press.
Scheines, R. [2005]: 'The Similarity of Causal Inference in Experimental and Non-experimental Studies', *Philosophy of Science*, **72**, pp. 927–40.
Schindler, S. [forthcoming]: 'Mechanistic Explanation: Asymmetry Lost', in V. Karakostas and D. Dieks (eds), *Recent Progress in Philosophy of Science: Perspectives and Foundational Problems*, Dordrecht: Springer.
Schurz, G. [2008]: 'Patterns of Abduction', *Synthese*, **164**, pp. 201–34.
Soom, P. [2011]: *From Psychology to Neuroscience. A New Reductive Account*, Heusenstamm: Ontos.
Soom, P. [2012]: 'Mechanisms, Determination and the Metaphysics of Neuroscience', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **43**, pp. 655–64.
Spirtes, P., Glymour, C. and Scheines, R. [2000]: *Causation, Prediction, and Search*, Cambridge: MIT Press.
Spohn, W. [2006]: 'Causation: An Alternative', *British Journal for the Philosophy of Science*, **57**, pp. 93–119.
Woodward, J. [2003]: *Making Things Happen*, Oxford: Oxford University Press.
Woodward, J. [2008]: 'Response to Strevens', *Philosophy and Phenomenological Research*, **LXXVII**, pp. 193–212.
Woodward, J. [2011]: 'Interventionism and Causal Exclusion', *PhilSci-Archive*, <http://philsci-archive.pitt.edu/8651/>
Zednik, C. [forthcoming]: 'Heuristics, Descriptions, and the Scope of Mechanistic Explanation', in C. Malaterre and P.-A. Braillard (eds), *How does Biology Explain? An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*, Dordrecht: Springer.