

Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms

Bert Leuridan

ABSTRACT

In this article, I present two conceptual problems for Craver's mutual manipulability account of constitutive relevance in mechanisms. First, constitutive relevance threatens to imply causal relevance despite Craver (and Bechtel)'s claim that they are strictly distinct. Second, if (as is intuitively appealing) parthood is defined in terms of spatio-temporal inclusion, then the mutual manipulability account is prone to counterexamples, as I show by a case of endosymbiosis. I also present a methodological problem (a case of experimental underdetermination) and formulate two partial, but fallible solutions based on the notions of parthood and synchronicity.

- 1 *Introduction*
 - 2 *Mechanisms*
 - 3 *Intralevel Causal Relations and Invariance*
 - 4 *Interlevel Constitutive Relations and Mutual Manipulability*
 - 5 *Doesn't Constitutive Relevance Entail Causal Relevance?*
 - 6 *Parthood and Endosymbiosis*
 - 7 *Possible Ways Out*
 - 8 *Interlevel Experiments*
 - 9 *The Problem of Underdetermination*
 - 10 *Solving the Problem of Underdetermination: Parthood*
 - 11 *Synchronicity, the Etiological Nature of Experimental Apparatus, and Causal-Constitutive Propagation*
 - 12 *Pragmatic Interests and the Boundaries of Mechanisms*
 - 13 *Concluding Remarks*
- Appendix*
-

1 Introduction

The concept of ‘complex-system mechanism’ plays an increasingly important role in philosophy of science (and in philosophy of the special sciences in particular). It is commonly defined such that a mechanism’s higher level behavior is realized by the organized behavior of its lower level parts (Glennan [1996], [2002]; Machamer et al. [2000]; Bechtel and Abrahamsen [2005]). This general characterization gives rise to two questions. First, how should these lower level behaviors be characterized? One of the most popular answers is that they can be described by Woodward-style invariant generalizations (Glennan [2002]; Woodward [2002]; Craver [2006], [2007]; Leuridan [2010]). Second, what is the ‘realization’ relation between the lower and the higher level?¹ A well-known answer is that this relation is constitutive, *not causal* (Craver and Bechtel [2007]).

In his book *Explaining the Brain*, Craver explicitly adopts Woodward’s account of causation ([2007], Chapter 3) and he proposes a ‘mutual manipulability’ account of constitutive relations (Chapter 4). This is perhaps the best and the most elaborate account of mechanisms available today. Yet in this article, I will show that it is problematic in several respects. In Sections 5–7, I present two main conceptual problems. Either constitutive relevance implies causal relevance (contra Craver and Bechtel [2007]) or we need a set of principled reasons why no cases of constitutive relevance are also cases of causal relevance. This I will call Problem #1 (Section 5). In my opinion, the best way to solve Problem #1 would be to focus on the notion of parthood. However, given a very natural definition of parthood (in terms of spatio-temporal inclusion), plenty of cases of bidirectional causation would unintentionally fit the mutual manipulability account of constitutive relevance, as I will illustrate by means of the endosymbiosis of aphids and *Buchnera*. This I will call Problem #2 (Section 6). In Section 7, I consider some possible ways out of this problem. In Sections 9–12, I discuss a methodological problem, the problem of underdetermination (which I will call Problem #3): it may be hard to distinguish between constitutive relevance and bidirectional causation on the basis of experimental evidence. I first present the problem and distinguish it from a related problem Craver discusses (Section 9). Then I present two partial, but fallible solutions to the problem of underdetermination, based on the notions of parthood (Section 10) and synchronicity (Section 11). Finally, I reject the suggestion that pragmatic considerations could help to solve the problem of underdetermination (Section 12). In Section 13, I briefly discuss the consequences of these three problems for those who try to characterize mechanisms in terms of Woodwardian invariant relations. The other sections

¹ I use ‘realization’ in a loose sense here, not in Jaegwon Kim’s sense (see Craver [2007], pp. 212–3).

serve to present the main concepts used: ‘Mechanisms’ (Section 2), ‘Invariance’ (Section 3), ‘Mutual manipulability’ (Section 4), and ‘Interlevel experiments’ (Section 8).

2 Mechanisms

Craver defines a mechanism as ‘a set of entities and activities organized such that they exhibit the phenomenon to be explained’ (Craver [2007], p. 5). He continues as follows: ‘Entities are the components or parts in mechanisms. They have properties that allow them to engage in a variety of activities. [...] Activities are the causal components in mechanisms. [...] Finally, the entities and activities in mechanisms are organized together spatially, temporally, causally, and hierarchically [...] The behavior of the mechanism as a whole requires the organization of its components [...]’ (Craver [2007], pp. 5–6).

For Craver, the explanandum phenomenon plays a central role in the above definition. ‘As Kauffman (1971) and Glennan (1996, 2002) argue, mechanisms are always mechanisms *of* a given phenomenon. The boundaries of mechanisms – what is in the mechanism and what is not – are fixed by reference to the phenomenon that the mechanism explains’ (Craver [2007], pp. 122–3, original emphasis). Craver occasionally calls the idea that all mechanisms are mechanisms *of* or *for* something, ‘Glennan’s law’; cf. (Glennan [2002], p. S344) for a statement of this view.

As is evident from this definition, a mechanism consists of two levels. At the micro-level, there are the mechanism’s entities and activities which are organized together spatially, temporally, and causally. At the macro-level, there is the phenomenon to be explained: the behavior of the mechanism as a whole.² What is important is that for Craver, all causal relations are intralevel. Interlevel relations, in contrast, are constitutive, *not causal*.

3 Intralevel Causal Relations and Invariance

How does Craver characterize a mechanism’s intralevel causal relations? Craver’s definition of ‘mechanism’ derives from Machamer, Darden, and Craver’s highly influential paper (MDC), which treats mechanisms as ‘entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions’ ([2000], p. 3). One of the central tenets of the MDC account is that it is the activities that account for the causing. But in an Anscombian vein, they refuse to give a general account of their causal nature: ‘Activities are types of causes. Terms like “cause” and

² The distinction between the micro- and the macro-level is not absolute. It is relative to the mechanism at hand. Levels of mechanisms are not monolithic divisions in nature (Craver [2007], p. 190). For a related view, see (Bechtel [2008], pp. 143–8).

“interact” are abstract terms that need to be specified with a type of activity and are often so specified in typical scientific discourse.’ (Machamer et al. [2000], p. 6)

Yet Craver’s views about causation have changed drastically since then. Since a couple of years ago, he understands the causal nature of activities in terms of Woodward’s manipulationist account (Craver [2006], p. 372).³ He describes this account as follows:⁴

‘The central idea is that causal relationships are distinctive in that they are potentially exploitable for the purposes of manipulation and control. More specifically, variable X is causally relevant to variable Y in conditions W if some *ideal intervention* on X in conditions W changes the value of Y (or the probability distribution over possible values of Y)’⁵ (Craver [2007], p. 94, emphasis added).

Craver defines *ideal interventions*, which I will call ‘ideal *etiological interventions*’ to contrast them with their ‘interlevel’ variants below, as follows.

An *ideal* [etiological] intervention I on X with respect to Y is a change in the value of X that changes Y , if at all, *only via* the change in X . More specifically, this requirement implies that:

- (I1) I does not change Y directly;
- (I2) I does not change the value of some causal intermediate S between X and Y except by changing the value of X ;
- (I3) I is not correlated with some other variable M that is a cause of Y ;⁶ and
- (I4) I acts as a ‘switch’ that controls the value of X irrespective of X ’s other causes, U . (Craver [2007], p. 96; original emphasis; adapted from Woodward and Hitchcock [2003])

(Whereas Craver treats I as an intervention, Woodward ([2003]) and Woodward and Hitchcock ([2003]) treat it as an intervention variable, some values of which give rise to an intervention. As I am convinced that this difference will play no role in my arguments, I will pass it by in the rest of this article. Another, much more important difference between Craver’s definition and theirs will show up in Section 5.)

³ Craver is not the only one to incorporate Woodward’s manipulationism in his theory of mechanisms. For related views, see (Glennan [2002]; Woodward [2002]; Leuridan [2010]).

⁴ The variables in question are random variables (determinables such as ‘sex’) that can take either finitely or infinitely many jointly exhaustive and mutually exclusive values (determinates such as ‘male’ and ‘female’).

⁵ Whereas Craver here formulates invariance under interventions merely as a sufficient condition for causation, Woodward treats it as a necessary and sufficient condition: ‘Invariance under at least one testing intervention (on variables figuring in the generalization) is necessary and sufficient for a generalization to represent a causal relationship or to figure in explanations’ (Woodward [2003], p. 250).

⁶ As it stands, (I3) is too restrictive. There is of course no reason to exclude I ’s being correlated (due to its being a cause of) with a causal intermediate M between I and X , provided M causes Y only via X ; see (Woodward [2003], p. 98).

A generalization is causal if (and only if?; see footnote 5) it is invariant under some range of interventions. This range of interventions need not be universal. Invariance under interventions is a gradual notion (Woodward [2003], pp. 257–65). As a result, invariant generalizations need not be universal or exceptionless (Woodward [2003], pp. 270–2). For Craver, this makes invariance the right concept to tackle causality and explanation in biology and in neuroscience. In his view, causal generalizations in biology rarely if ever satisfy the traditional criteria for lawfulness (necessity, universality, etc.): they are limited in scope, stochastic, mechanistically fragile, and historically contingent (Craver [2007], pp. 66–9). Yet they can function in explanation. Woodward’s manipulationist account explains why and how they do. Moreover, Woodward’s account also nicely fits (etiological) experimental practice in neuroscience (Craver [2007], pp. 93–8).

4 Interlevel Constitutive Relations and Mutual Manipulability

Intralevel relations are causal in Craver’s view if they are invariant under some range of ideal etiological interventions. But how should we conceive of the relation *between* the levels?

Craver and Bechtel’s ([2007]) answer is negative: interlevel relations are constitutive, not causal.⁷ They claim not only that ‘constitutive relation’ and ‘causal relation’ are distinct concepts, but also that *no* constitutive relation *can* also be a causal relation. They have three arguments (see Craver [2007], pp. 153–4; Craver and Bechtel [2007], pp. 552–4). These will play an important role in this article.

- Symmetry argument: constitutive relations are symmetric⁸ whereas many, if not most, causal relationships are unidirectional.
- Logical argument: ‘[. . .] in the constitutive relation, a token instance of the property ψ is, in part, constituted by an instance of the property ϕ ; as such, the tokening of ϕ is not logically independent of the tokening of ψ . At least since Hume, many philosophers have held that causes and effects must be logically independent’ (Craver [2007], p. 153).
- Synchronicity argument: interlevel relations are synchronic whereas on many accounts of causation causes must precede their effects.⁹

⁷ The main argument of their paper is that there is no interlevel causation. ‘We argue that intelligible appeals to interlevel causes (top-down and bottom-up) can be understood, without remainder, as appeals to mechanistically mediated effects. Mechanistically mediated effects are hybrids of causal and constitutive relations, where the causal relations are exclusively intralevel’ (Craver and Bechtel [2007], p. 547).

⁸ For a criticism of this claim, see (Kistler [2009], p. 601).

⁹ ‘Synchronicity’ is not used in the Jungian sense here.

Craver ([2007]) offers a positive answer to the question how we should conceive of the interlevel relations by providing an account of *constitutive relevance*. His working account is as follows:¹⁰

[A] component is relevant to the behavior of a mechanism as a whole when one can wiggle the behavior of the whole by wiggling the behavior of the component *and* one can wiggle the behavior of the component by wiggling the behavior as a whole. The two are related as part to whole and they are *mutually manipulable*.¹¹ (Craver [2007], p. 153, original emphasis)

This working account is explicated as follows:

More formally: (i) X is part of S ; (ii) in the conditions relevant to the request for explanation there is some change to X 's ϕ -ing that changes S 's ψ -ing; and (iii) in the conditions relevant to the request for explanation there is some change to S 's ψ -ing that changes X 's ϕ -ing. (Craver [2007], p. 153)

The conditions (ii) and (iii) are then further explicated in terms of Woodward's manipulability account (or a revised version thereof, since constitutive relations are said to be non-causal) by means of the following four principles, which together with the parthood condition constitute the *mutual manipulability account* of constitutive relevance:

(CR1) When [in the conditions relevant to the request for explanation] ϕ is set to the value ϕ_1 in an ideal intervention, then ψ takes on the value $f(\phi_1)$. (Craver [2007], p. 155)

[Moreover] there should be an ideal intervention on X 's ϕ -ing that changes the value of S 's ψ -ing under the conditions (CR1a) that the intervention, I , leaves all of the other dependency relations in S 's ψ -ing unchanged and (CR1b) that other interventions have removed the contributions of other redundant components. (Craver [2007], pp. 156–7)

(CR2) [I]f [in the conditions relevant to the request for explanation] ψ is set to the value ψ_1 in an ideal intervention, then ϕ takes on the value $f(\psi_1)$. (Craver [2007], p. 159)

This notion of ideal intervention is not the one we encountered in Section 3. In the context of constitutive relevance, ideal interventions are characterized as follows (let me call them 'ideal *interlevel* interventions'):

[A]n *ideal* [interlevel] intervention I on ϕ with respect to ψ is a change in the value of ϕ that changes ψ , if at all, *only via* the change in ϕ . This implies that:¹²

(II_c) the intervention I does not change ψ directly;

¹⁰ Craver ([2007], p. 141) explicitly intends this to be a sufficient condition for constitutive relevance.

¹¹ The predicate 'component' applies both to entities and activities (Craver [2007], p. 7). Note that components should not be confused with mere parts.

¹² Note the similarity with conditions (I1)–(I4). This similarity will play an important role in the following section.

- (I2_c) *I* does not change the value of some other variable ϕ^* that changes the value of ψ except via the change introduced into ϕ ;
 (I3_c) *I* is not correlated with some other variable *M* that is causally independent of *I* and also a cause of ψ ; and
 (I4_c) *I* fixes the value of ϕ in such a way as to screen off the contribution of ϕ 's other causes to the value of ϕ . (Craver [2007], p. 154, original emphasis)

This account of interlevel relations raises two questions. The first question, to be discussed in the following section, concerns the relation between constitutive relevance and causal relevance. Doesn't the former entail the latter? The second question concerns the plausibility of (CR1a). This problem is rather unrelated to my main argument; but I think it is interesting in its own right, so let me tackle it here.

Craver supplements (CR1) with the principles (CR1a) and (CR1b) because (CR1) does not provide a necessary condition for *X*'s ϕ -ing to be relevant to *S*'s ψ -ing. 'Satisfying CR1 is unnecessary because compensatory responses (such as recovery, redundancy, and reorganization) can prevent changes to *S*'s ψ -ing [...]' (Craver [2007], p. 156). Recovery, redundancy, and reorganization are actual problems faced by experimental researchers in neuroscience and in the life sciences in general. (CR1a) excludes cases of recovery and reorganization. (CR1b) deals with cases of redundancy. Whereas the latter amendment is fairly acceptable (see footnote 13), in a sense the former is not. Recovery and reorganization (also known as robustness) are both common and important phenomena: 'The capacity for robustness appears to be both ubiquitous and significant for biological systems. Indeed robustness gives evolved complex systems, including organisms and ecosystems, the ability to survive in an environment with changing internal and external conditions' (Mitchell [2009], pp. 71–2). The question whether (CR1a) is plausible is thus important.¹³

One may defend (CR1a) by stressing that it is fairly weak, due to the existential quantification over interventions. Craver merely demands that there is *an* ideal interlevel intervention that does not elicit reorganization (let me call this a neutral intervention), whereas robustness only entails that some (or perhaps most) interventions are not neutral. The assumption that for some robust system there is a neutral intervention is thus methodologically

¹³ That (CR1b) is fairly acceptable can be seen from Mitchell's work on complexity. In the context of etiological research on complex systems (she focuses on gene expression), the absence of changes in *Y* given a single intervention on *X* is *not* considered a sufficient reason to conclude that *X* is *not* a cause of *Y*. Complex systems are investigated by means of multivariate experimental techniques and network analysis. 'Exploring multiple scenarios rather than shielded experiments on targeted single components may prove more conducive to understanding context-rich causal structures' (Mitchell [2009], p. 83). This use of multivariate techniques is what (CR1b) comes down to.

acceptable. Yet arguably it is metaphysically wrong.¹⁴ Even if most robust systems are such that there is a neutral intervention, there is no reason to assume this holds for *all* robust systems. Given that (CR1a) is meant to take part in the definition of constitutive relevance, it is too strong.

5 Doesn't Constitutive Relevance Entail Causal Relevance?

The above characterization of constitutive relevance arouses suspicion. It looks as if constitutive relevance implies causal relevance, even though Craver ([2007]) and Craver and Bechtel ([2007]) explicitly deny this. If X 's ϕ -ing is constitutively relevant for S 's ψ -ing, then changing X 's ϕ -ing is a good means to change or control S 's ψ -ing. But then, isn't X 's ϕ -ing causally relevant for S 's ψ -ing? Let us have a look in more detail.

Suppose that X 's ϕ -ing is constitutively relevant for S 's ψ -ing. Then there is an ideal interlevel intervention I on ϕ with respect to ψ such that the intervention on ϕ changes ψ , if at all, only via the change in ϕ . Since, by hypothesis, I is an ideal interlevel intervention, conditions (I1_c) to (I4_c) are satisfied. As I will now show, it follows that conditions (I1)–(I4) all are satisfied too. The question then is whether this implies that I is an ideal etiological intervention on ϕ with respect to ψ .¹⁵ If I is an ideal etiological intervention on ϕ with respect to ψ , then X 's ϕ -ing is causally relevant to S 's ψ -ing.

First, (I1) is satisfied because (I1_c) is—just substitute ϕ and ψ (which are variables the values of which denote activities) for X and Y (which arguably play the same role in the definition of ideal etiological interventions), respectively. Second, by (I2_c) I does not change the value of any other variable ϕ^* that changes the value of ψ except via the change introduced into ϕ . Hence, *a fortiori* it does not change the value of any causal intermediate, if any, between ϕ and ψ except by changing the value of ϕ ; so (I2) is satisfied too. Third, condition (I3) seems the hardest part, but it is satisfied as well. By (I3_c), I is not correlated with any other variable M that is causally independent of

¹⁴ This case is similar to the status of faithfulness in causal discovery. Faithfulness is an assumption relating to causal graphs and probability distributions that greatly simplifies algorithms for causal discovery (Spirtes et al. [2000], pp. 13–4). A probability distribution $P(V)$ and a directed acyclic graph $G = \langle V, E \rangle$ satisfy the faithfulness condition if and only if every probabilistic independence relation true in P is entailed by the Markov condition as applied to G , i.e. P verifies no 'extra' independence relations. (The Markov condition applied to G states for each node in V that, conditionally on its graphic parents, it is probabilistically independent of its graphic non-descendants.) The faithfulness assumption is methodologically acceptable, yet metaphysically wrong. It is methodologically acceptable since the probability of randomly drawing an unfaithful distribution from the set of distributions that are causally Markov to G , is zero; see (Meek [1995]) for a more rigorous statement in terms of Lebesgue measure. It is metaphysically wrong because unfaithful distributions do exist.

¹⁵ In the following exposition I will *not* assume that ϕ and ψ are *not* causally related, because that is precisely what is at stake. Therefore, I will assume it makes sense to talk about the causal intermediates between ϕ and ψ .

I and also a cause of ψ . In condition (I3), no restriction to causes of M that are causally independent of I is made. So what about cases in which I is correlated with some M such that I is a cause of M and M is a cause of ψ ? These are dealt with by (I2_c). Either I changes the value of M (which in turn changes the value of ψ) only via ϕ , or it does not. The first case is allowed for by (I2_c) and should not be considered a counterexample to (I3); the second case is excluded by (I2_c). Fourth, (I4) is satisfied because (I4_c) is (by substitution).

If I satisfies (I1_c)–(I4_c), then it satisfies (I1)–(I4). The question now is whether this shows that I is an ideal etiological intervention on ϕ with respect to ψ . If it were, X 's ϕ -ing would be causally relevant for S 's ψ -ing.

There is some room for interpretation here. Craver's definition of ideal etiological interventions (see Section 3) differs both from Woodward and Hitchcock's definition ([2003], pp. 12–3) and from Woodward's definition ([2003], p. 98). Craver formulates his clauses (I1)–(I4) as implications (and thus, it seems, as *necessary* conditions) of his general characterization of ideal etiological interventions. Woodward and Hitchcock treat similar (but slightly different) conditions as *sufficient* for some variable I 's being an intervention variable. Woodward treats similar (but again slightly different) conditions as both *necessary* and *sufficient* for some I 's being an intervention variable. (Woodward ([2003]) and Woodward and Hitchcock ([2003]) do not explain the difference between their definitions; Craver ([2007]) does not explain the difference between his definition and those of Woodward ([2003]) or Woodward and Hitchcock ([2003])). If we follow either Woodward's or Woodward and Hitchcock's characterization, the above analysis counts as a genuine proof of the thesis that X 's ϕ -ing being constitutively relevant for S 's ψ -ing implies its being causally relevant for S 's ψ -ing as well. If it can be shown that (clauses similar to) (I1)–(I4) can be deduced from (I1_c)–(I4_c), it is proved that I is an ideal etiological intervention.¹⁶

If we follow Craver's definition, ' X 's ϕ -ing is causally relevant for S 's ψ -ing' does not follow deductively from ' X 's ϕ -ing is constitutively relevant for S 's ψ -ing'. But then the question is: which are the cases such that I satisfies (I1)–(I4), whereas X 's ϕ -ing is not causally relevant for S 's ψ -ing? To safeguard his thesis that *no* constitutive relations are at the same time also causal relations, Craver needs to show that for *all* cases in which I satisfies (I1)–(I4) and also (I1_c)–(I4_c), I is not a genuine ideal etiological intervention. Of course, the reason he adduces may not be *ad hoc*.

Craver's current position (personal communication) is that criteria such as (I1)–(I4) are neither necessary nor sufficient for I 's being an ideal etiological intervention. They are not sufficient because they do not include a modularity

¹⁶ In the appendix, I show that Woodward's clauses (I1_w)–(I4_w), which he considers to be necessary and sufficient for I 's being an ideal etiological intervention, can be derived from Craver's definition of ideal interlevel interventions.

requirement. He now tends to treat them as potent, yet fallible, indicators of causal relevance relations obtaining among a system of variables. They are fallible because, e.g. modularity may fail (for some system) (end of personal communication).

Before I assess this argument, let me briefly discuss modularity. Strictly speaking, modularity is a feature of systems of structural equations (or more generally of causal models):

It says that each structural equation in a system of structural equations that correctly captures the causal relations among a set of variables is invariant under interventions that disrupt other equations in the system by setting the values of their dependent variables within some limited range¹⁷ (Hausman and Woodward [1999], pp. 542–3).

Derivatively, modularity is a feature of that part of reality that is correctly described by the modular system of equations (or the causal model): '[. . .] it is nature and in particular facts about what happens or would happen under interventions that determine whether a given system of equations is modular' (Hausman and Woodward [1999], p. 544). As such, it states that distinct causal relations are independently disruptable: 'The central presupposition underlying this discussion of modularity [. . .] is that if two mechanisms¹⁸ are genuinely distinct it ought to be possible (in principle) to interfere with one without changing the other' (Hausman and Woodward [1999], p. 549).

I acknowledge that modularity may fail, especially in such 'messy' causal systems as studied by neuroscience or by the biological and the social sciences in general (cf. my discussion of (CR1a) and robustness at the end of the previous section). Yet there are three problems with using failure of modularity to block the inference from constitutive to causal relevance.¹⁹ First, failure of modularity has little to do with Craver and Bechtel's reasons to deny interlevel relations a causal interpretation. Second, it is doubtful that *all* cases in which *I* satisfies *both* (I1)–(I4) and (I1_c)–(I4_c) are cases in which modularity fails. Hence, even if failure of modularity would be a good reason to adduce in some of the cases, it should be backed by other reasons to cover the remaining cases. Third, this solution is inconsistent with the mutual manipulability account of constitutive relevance. Modularity, like invariance, is a gradual notion: '[. . .] when one intervenes and sets the value of *X* (thereby disrupting the equation (mechanism) that determined the value of *X*), the

¹⁷ This definition of modularity has been heavily disputed: see (Cartwright [2002], p. 417; Hausman and Woodward [2004], p. 149; Steel [2006], p. 228) for alternative definitions.

¹⁸ 'Mechanisms' is not used here in the sense of Section 2, but as the ensemble of an effect and its causes; so the definition is wholly etiological.

¹⁹ The personal communication mentioned above only addressed the question whether (I1)–(I4) are necessary and/or sufficient for *I*'s being an ideal etiological intervention, not whether the answer to this question could be used to block the inference from constitutive to causal relevance.

other equations will remain invariant *only for some* range of values of X (Hausman and Woodward [1999], p. 545, emphasis added). So strictly speaking failure of modularity only occurs if the other equations/mechanisms change for *all* values of X . But this is ruled out by (CR1a).

As it stands, either constitutive relevance implies causal relevance (if (I1)–(I4) are treated as sufficient conditions) or we need a set of principled reasons why ideal interlevel interventions, all of which satisfy (I1)–(I4), cannot be ideal etiological interventions. Let me call this Problem #1.

Can this problem be solved without abandoning the strict distinction between constitutive and causal relevance? In my opinion, the best way would be to focus on the parthood condition in the mutual manipulability account of constitutive relevance, because seemingly that would require no changes to Craver's views. Craver and Bechtel's logical argument is that causes and their effects must be distinct events, objects, or processes. Since mechanisms and their components are related as wholes and parts, they are not genuinely distinct. Invoking parthood would not be *ad hoc* and it would cover all cases of constitutive relevance at once. The argument can be turned into a straightforward solution to Problem #1. Consider the following assumption:

Assumption 1: if X is part of S , then an intervention I on X directly changes S .

If we accept Assumption 1, then we must conclude that condition (I1) is violated, that the intervention in question is not an ideal etiological intervention, and that X and S are not causally related. Unfortunately, this solution has two important drawbacks. First, if we accept Assumption 1, it is hard to deny the following:

Assumption 2: if X is part of S , then an intervention I on X 's ϕ -ing directly changes S 's ψ -ing.

If we accept Assumption 2, then we must conclude that condition (I1_c) is violated too, that the intervention in question is not an ideal interlevel intervention either, and that X and S are also not constitutively related. Thus the above solution would only work if Craver revised his concept of ideal interlevel intervention.

The second drawback is that it requires that we have a good account of the parthood relation. As I will show in the following section, it is hard to find an account of parthood that does not cause further problems for the mutual manipulability account of constitutive relevance.

6 Parthood and Endosymbiosis

In the previous sections we saw that constitutive relevance is defined partly in terms of parthood and that this may sow the seeds of a solution to Problem

#1. But when is X part of S ? It will emerge that there is no straightforward answer to this question in the present context.

Formal mereology is not of any help here. It mostly treats (proper) parthood as a primitive relation having certain formal properties (in standard mereology, proper parthood is transitive, asymmetric, and irreflexive), but it gives no criterion for parthood itself. ‘The relation of proper part, $<$, does not require much illustration, since it is firmly embedded in our ordinary way of conceptualizing the world; it holds, for example, between a man and his forearm’ (Koslicki [2008], p. 11).

In my view, the most natural definition is in terms of spatio-temporal inclusion:

(STI) X is part of S iff the spatio-temporal region occupied by X is contained in the spatio-temporal region occupied by S .²⁰

Although STI is intuitively appealing, it leads to problems in the context of the mutual manipulability account.²¹ Consider the following definitions:

Symbioses are long-term associations between different species typically with a substantially smaller *symbiont* living in or on a considerably larger multicellular *host*. Microbes are often *endosymbionts* living inside of host bodies. [...] *Intracellular* symbionts (e.g. bacteria) are typically found in the cytoplasm of host cells. [...] In a *mutualistic symbiosis*, both symbiont and host benefit. In an *obligate mutualism*, both symbiont and host require the presence of the other. (Lombardo [2008], p. 480, original italics)

A well-known example of obligate mutualistic endosymbiosis is provided by *Buchnera* bacteria and aphids. Aphids, like many other insect hosts, have bacteriocytes (also known as ‘mycetocytes’). These are specialized cells that host bacterial symbionts (Ishikawa [2003], p. 6). One of the most important symbionts they host is *Buchnera*. *Buchnera* and aphids are mutually dependent for their nutrition (for the experimental evidence underlying these claims, see Section 9).

One of the most characteristic features of *Buchnera*, unveiled by our genome analysis, is that the genes for biosyntheses [*sic*] of the amino acids essential for the aphid hosts [...] are present, but those for the non-essential amino acids are almost completely missing [...]. This complementarity of the gene repertoire shows how successfully the

²⁰ STI is restricted to parts and wholes that have a spatio-temporal location. Thus, ‘Rationality is part of personhood’ may be true, even though rationality and personhood do not satisfy STI. (See Varzi [2009] for examples of parts and wholes that have no spatio-temporal location.) STI is similar in spirit to Goodman’s account of parthood in terms of ‘overlap’ ([1977], p. 35) and it equally treats of parthood as a reflexive relation.

²¹ Craver ([2007], pp. 141–2, 187–8) rejects spatial containment as a criterion for componenty, yet I am focussing on the broader category of parthood; ‘Not all parts are components’ (Craver [2007], p. 140).

symbiosis is operating, in that *Buchnera* provides the host with what the host cannot synthesize, and conversely, the host provides the symbiont with what *Buchnera* cannot synthesize. Moreover, as the precursors of some essential amino acids are non-essential amino acids, glutamate and aspartate [...], the biosynthetic pathways of both the host and the symbiont are not only complementary, but also mutually dependent. This analysis is consistent with experimental evidence that aphids do not usually excrete a nitrogenous waste product, but recycle the amino groups as glutamine, which *Buchnera* uses as a substrate for the synthesis of essential amino acids [...]. (Shigenobu *et al.* [2000], p. 82)

An amino acid is essential for an organism (*in casu* the aphids) if the organism cannot synthesize it in sufficient quantities; it must therefore be present in the diet (Martin and Hine [2008]). But the aphids' natural diet, phloem sap, is too poor to provide the essential amino acids in large enough amounts (Douglas [1993]). The essential amino acids in question are synthesized by *Buchnera*. The *Buchnera* in turn need amino acids such as glutamine, aspartate, and glutamate that are non-essential for the aphids, but which they cannot synthesize themselves and hence must receive from their hosts.

If we accept both the mutual manipulability account of constitutive relevance and **STI** as a definition of parthood, the *Buchnera* should be considered components of their hosts, since the following hold:

- (i) *Buchnera* are spatio-temporally contained in the aphids' bacteriocytes, and thus are part of their hosts;
- (ii) in the conditions relevant to the request for explanation (*viz.* when aphids feed on phloem sap), there is some change to the *Buchnera*'s φ -ing that changes the aphid's ψ -ing; and
- (iii) in the conditions relevant to the request for explanation, there is some change to the aphid's ψ -ing that changes the *Buchnera*'s φ -ing.

In view of Glennan's law that every mechanism is *of* or *for* something (see Section 2), we must delineate the explanandum activity (ψ) fairly precisely. Likewise, we must clearly delineate the purported component's relevant activity (φ). As is evident from Shigenobu *et al.* ([2000], p. 82; see also their figures 2 and 3), ψ is the synthesizing of amino acids that are non-essential for the aphids; φ is the synthesizing of amino acids that are essential for the aphids.²²

²² Of course, there is more than one essential amino acid that plays a role in this symbiotic relationship, just like there is more than one non-essential amino acid. So strictly speaking, we should distinguish different explanandum activities (ψ_1 -ing, ψ_2 -ing, ...) and different component activities (φ_1 -ing, φ_2 -ing, ...). Which ψ_i 's are related to (or needed for) which φ_j 's, and vice versa, is represented by biosynthetic pathways. So the *Buchnera* and the aphids bear seemingly constitutive relationships in multiple ways. But that does not conflict with Glennan's law (each of the φ_i and ψ_j are well delineated) and by itself is no reason to say these relations cannot be constitutive.

Even though the aphid–*Buchnera* relationship nicely fits Craver’s working account, there are good reasons why *Buchnera* should not be considered components. First and foremost, the relation between the aphid’s and the *Buchnera*’s respective activities is diachronic. The production of each token amino acid by the *Buchnera*, precedes the use of that token amino acid in the aphid’s synthesizing (a token instance of) another type of amino acid, *and vice versa*. Constitutive relations, in contrast, are said to be synchronic (see Section 4 for Craver and Bechtel’s criterion, and Section 11 for a methodological guideline based on this criterion). Second, hosts and their symbionts are by definition members of distinct species (cf. the definitions at the beginning of this section) and thus must themselves also be distinct (whereas mechanisms and their components are not distinct). Third, whereas endosymbionts are located in their hosts, ectosymbionts are not. Therefore, they should not be considered **STI**-parts (and *a fortiori* components) of their hosts. But this differential treatment of endosymbionts and ectosymbionts would be rather arbitrary.

(This is not to say that *Buchnera* and aphids—or hosts and their endo- or ectosymbionts in general—cannot figure as components in the same (ecological) mechanism. But still the relation *between* the symbionts is causal, not constitutive.²³ This is also not to say that *Buchnera* or aphids cannot themselves be treated as ‘wholes’ for mechanistic decomposition, only that in a decomposition of the aphids, the *Buchnera* cannot serve as components.)

The case of *Buchnera* and aphids is not a rarity. Plenty of similar counter-examples can be found. Mutualistic endosymbiosis is widespread and has played an important role in the evolution of many species:

Today, the world abounds with mutualistic symbioses between microbes and multicellular hosts, and hundreds of examples can be found [...]. It is probable that all multicellular organisms will be found to exist in obligate mutualistic symbiotic relationships with microbes that aid them in nutrition and defense against predators or pathogens [...] (Lombardo [2008], p. 482).

In short, if **STI** were accepted as an account of parthood, plenty of cases of bidirectional causation would unintentionally fit the mutual manipulability account of constitutive relevance. Thus the mutual manipulability account would not provide a sufficient condition for constitutive relevance (contra Craver’s claim it does provide a sufficient condition; [2007], p. 141). Let me call this Problem #2.

A corollary of Problem #2 is that one cannot solve Problem #1 by simply restricting the scope of application of the manipulationist account of causation (as a semantic or interpretive project, cf. (Woodward [2003], p. 95)

²³ For an account of ecological mechanisms which integrates Woodward’s account of causation, see (Păslaru [2009]).

so that it cannot be applied to things which are related as parts and wholes. For that would presuppose one has a clear account of ‘parthood’.²⁴

7 Possible Ways Out

How could one solve Problem #1 without abandoning the strict distinction between constitutive and causal relevance and without inviting Problem #2? One solution would be to adopt another definition of parthood. However, I do not see which definition would do—I consider **STI** the most natural definition. Therefore, I will not try to provide an alternative definition of parthood here, but merely sketch three obvious constraints for any such alternative. First, on pain of circularity the definition should not rely on the notion of constitutive relevance. Second, the definition should be such that parthood does not coincide with componency. Otherwise, clause (i) of the working account would be superfluous and the distinction between mere parts and relevant parts would vanish. These were part of the reasons why I opted for **STI**: spatio-temporal containment does not coincide with componency, and it is devoid of any appeal to constitutive relevance. Third, one may not just drop the requirement that X be part of S from the mutual manipulability account. Otherwise, all cases of circular or bidirectional causation (such as causal feedback mechanisms, cf. Section 9) would count as cases of constitutive relevance.

Another solution would be to add another requirement to the mutual manipulability account, to wit synchronicity. Not only should X be part of S and should X 's ϕ -ing and S 's ψ -ing be mutually manipulable. For X 's ϕ -ing to be constitutively relevant for S 's ψ -ing, the two should also be synchronic. In my view, to spell this out would be a very promising move to make on Craver's behalf. However, let me point to three further considerations. First, it only makes sense to say that *instances of X 's ϕ -ing* are synchronic with *instances of S 's ψ -ing*. If X 's ϕ -ing and S 's ψ -ing are treated as type-level variables, they are neither synchronic nor diachronic. Second, there is no guarantee that adding this further condition would make the mutual manipulability account immune to exceptions. If causation is defined along interventionist lines, then there exist cases of synchronic causation (see Section 11). So it is at least conceivable that there are cases of synchronic mutual manipulability that satisfy the parthood condition (as defined by **STI**). Of course, unless one finds a convincing example, Craver need not worry too much about this mere possibility. Third, in the rest of this article, I will discuss more practical problems relating to mechanisms and constitutive relevance. Even though it is relatively clear what it *means* to say that token instances of two variables are synchronic or

²⁴ This does not preclude using *knowledge* of part-whole relations to solve *methodological* problems in *practice* (see Section 10).

diachronic, it may not be so easy to distinguish between synchronic and diachronic relations *in practice* (see Section 11).

8 Interlevel Experiments

Up until now, I have been focusing on conceptual (or if you wish, metaphysical) issues. What are intralevel causation and interlevel constitutive relevance? Are the two really distinct? What is parthood? ... From now on, I will focus on epistemological or methodological questions, such as how we can find out that X 's ϕ -ing is constitutively relevant for S 's ψ -ing.

Craver's mutual manipulability account of constitutive relevance is derived from actual 'interlevel' experimental practice in neuroscience and the biomedical sciences in general. Interlevel experiments are like experiments for testing etiological causal claims, but there is one important difference. In the latter, the intervention and detection techniques are applied to the same mechanistic level. In the former, they are applied to different levels of mechanisms (Craver [2007], p. 145).

There are different kinds of interlevel experiments (Craver [2007], pp. 146–52; Craver and Bechtel [2007], pp. 553–4). Interlevel experiments can either be *bottom-up* or *top-down*. In the first case, the intervention techniques are applied to the level of the mechanism's purported components and the detection techniques are applied to the level of the explanandum phenomenon. In the latter case, this is the reverse. Interlevel experiments can also be either *inhibitory* (when the entities and activities intervened on are inhibited, diminished, disabled, destroyed, etc.) or *excitatory* (when they are stimulated, excited, enhanced, etc.).

These distinctions jointly give rise to four kinds of experiments. Interference experiments are bottom-up and inhibitory. Stimulation experiments are bottom-up and excitatory. Activation experiments are top-down and excitatory.²⁵ The last kind, viz. top-down inhibitory deprivation experiments, is quite rare in neuroscience. Each of these kinds faces certain inferential challenges which can partly be remedied by performing them in tandem, i.e. by combining bottom-up and top-down experiments. This is part of the motivation for Craver's *mutual* manipulability account ([2007], p. 158).

In my view, Craver's endeavor to tie his account of constitutive relevance to actual interlevel experimental practice is highly commendable. By analyzing what interlevel experiments are, what inferential challenges they face, and how these challenges can be remedied, his account of mechanisms may not only be interesting for the philosopher of science, but also for the practising scientist. In the remaining sections, I would like to add to this undertaking by pointing

²⁵ In (Craver [2002]), these were called additive experiments.

to an inferential challenge that Craver hitherto has not addressed and by suggesting some partial (and fallible) remedies.

9 The Problem of Underdetermination

In Section 6, we have seen that cases of mutualistic endosymbiosis generate a problem (Problem #2) for the mutual manipulability account of constitutive relevance. If parthood is defined along the lines of **STI**, the mutual manipulability account does not provide a sufficient condition for constitutive relevance. Problem #2 has a methodological counterpart. Consider the following quote:

Aphid requirement for the symbiosis [with *Buchnera*] has been demonstrated by studies of bacteria-free aphids, usually called aposymbiotic aphids. These aphids can be generated experimentally by antibiotic treatment, usually chlortetracycline or rifampicin administered either orally or by injection into the hemolymph [...]. At the appropriate dosage, these antibiotics have minimal direct deleterious effect on insect metabolism or behavior [...]. (Douglas [1998], p. 21)

Experiments with aposymbiotic aphids have been crucial in the study of the aphid–*Buchnera* symbiosis. However, they strongly resemble interference experiments. First, they are obviously inhibitory (the bacteria are removed in order to inhibit their ϕ -ing). Second, since the bacteria are **STI**-parts of the aphids, one could be tempted to regard the experiment as bottom-up rather than etiological. Third, the researchers make sure that the aphids' ψ -ing is not directly changed, cf. condition (II_c) above. Fourth, in line with the idea that mechanisms are always mechanisms *of* or *for* a given phenomenon (see Glennan's law in Section 2), the experiments are aimed at specific explanandum activities or functions ψ (and component activities φ). *Buchnera* contribute to the aphids' essential amino acid requirements, but, most probably, not to their lipid, sterol, or vitamin requirements (Douglas [1998], pp. 27–9). Conversely, the aphids contribute to specific needs of the *Buchnera*, by producing non-essential amino acids, some of which the *Buchnera* need as precursors to produce essential amino acids (Shigenobu *et al.* [2000], p. 82).²⁶

Experiments on aphids and *Buchnera* resembling top-down interlevel experiments are less common, due to several methodological obstacles. Still, some can be compared with deprivation experiments:

Preparations of isolated *Buchnera* can be obtained by techniques akin to the protocols for the isolation of organelles, such as mitochondria [...]. The isolated *Buchnera* preparations are viable and metabolically active for several hours, and they can be used to establish basic metabolic

²⁶ See Section 6, where I discussed the precise delineation of ψ and φ .

capabilities, such as capacity for aerobic respiration and DNA and protein synthesis [...]. These *Buchnera* preparations are not, however, metabolically equivalent to *Buchnera* in symbiosis, as is illustrated by the dramatic effect of isolation on the array of proteins synthesized by the bacteria [...]. (Douglas [1998], p. 22)

Isolation of *Buchnera* comes down to depriving them from the aphid's ψ -ing. The result is a change (a dramatic effect) in their ϕ -ing.

Experiments aimed at the study of metabolic versatility of *Buchnera* in turn resemble activation and/or deprivation experiments. The metabolic versatility of *Buchnera* is the extent to which the nutrients they provide to their aphid host may vary, in accordance with the aphid's demand for those nutrients (Douglas [1998], p. 30). Consider the following experiment:

When a single clone of *A. pisum* [Pea aphid] was reared on chemically defined diets of different amino acid composition, the free amino acid titers of symbiotic aphids were very uniform, whereas those of aposymbiotic aphids bore distinct similarities to the amino acid composition of the diets on which they were feeding [...]. These data were interpreted as evidence for a breakdown in regulation of the free amino acid pools in aposymbiotic aphids. The implication is that the bacteria contribute directly to amino acid homeostasis in symbiotic aphids, i.e. control over the profile of amino acids provided by the bacteria is so fine-tuned that it maintains the optimal amino acid titers of the aphid body fluids. (Douglas [1998], p. 31)

In what sense does this experiment resemble activation or deprivation experiments? The manipulation of the aphid's diet resembles top-down interventions. Depending on the chemical composition of the diets in the respective experimental groups, this manipulation is either inhibitory or excitatory. The difference between the free amino acid titers of symbiotic and aposymbiotic aphids provides evidence for *Buchnera*'s adaptive ϕ -ing in symbiotic aphids and hence may count as a detection technique at the bottom-level.

In short, research into aphid-*Buchnera* endosymbiosis is to a large extent based on experiments that closely resemble interlevel experiments. They are inhibitory or excitatory, they appear 'top-down' or 'bottom-up' and they are performed in tandem with other, 'reverse' experiments (i.e. experiments that appear 'bottom-up' or 'top-down').

This situation is not confined to the study of endosymbiosis. Plenty of etiological relations are symmetrical or bidirectional. Examples of causal feedback are legion in the biological and the social sciences and they are certainly not rare in neuroscience.²⁷ Even more importantly, causal feedback is

²⁷ Bibliographic search in the ISI Web of Knowledge for 'topic=(causal feedback)' on 25 August 2009 yielded 1302 hits, 214 of which are in psychology, 210 in the behavioral sciences, 94 in the neurosciences and neurology, 55 in biochemistry and molecular biology—these are the scientific fields that Craver intends to address. One year later (on 24 September 2010) these numbers have

considered one of the motors generating biological complexity and higher level phenomena—the very phenomena Craver wants to explain (Mitchell [2003], pp. 38–57, [2009], pp. 34–44).²⁸ With these bidirectional causal relations go experiments that treat a variable one time as a cause variable, another time as an effect variable.²⁹

The question, or the methodological problem, now is: how do we know these experiments are not top-down or bottom-up, but merely etiological? Let me call this the ‘problem of underdetermination’ or, simply, Problem #3. The issue seems insoluble when we allow ourselves to use only the resources of the mutual manipulability account. Therefore, I will discuss some additional resources in the next two sections. Craver underestimates this problem:

In sum, I conjecture that to establish that X 's ϕ -ing is [constitutively] 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. (Craver [2007], p. 159, emphasis added)

Craver does address a distinct problem of underdetermination and he proposes some possible solutions. The problem he addresses is that the distinction between a component and a background condition is likely to be vague (Craver [2007], p. 143). For example, activation experiments may suffer from an inferential challenge he calls *mere correlates* or sterile effects. ‘[...] intervening to make S ψ might activate some component X of S , but the activation of that component has sterile effects, relative to S , on some irrelevant part, C . C would then be strongly correlated with task activation, but it would not be part of the mechanism’ (Craver [2007], p. 152). C should then be treated as a ‘poorly understood background condition’. Likewise, interference experiments may suffer from an inferential challenge he calls *indirect interference*: interference with X 's ϕ -ing may disrupt S 's ψ -ing even if the former is

even dramatically increased, especially in Craver's fields of interest: 1456 hits in total (+11%), 254 in psychology (+19%), 247 in the behavioral sciences (+18%), 114 in the neurosciences and neurology (+21%), 73 in biochemistry and molecular biology (+33%). (I must admit I do not know what proportion of this increase is due to the inclusion of new journals in the ISI Web of Knowledge.) It is also interesting to note that Craver's own schematic representation of mechanisms includes a causal cycle at the lower level (Craver [2007], p. 7).

²⁸ Craver admits that symmetric causal relations exist, but he seems to underrate both their incidence and their importance: ‘Although there are *some* cases of cause and effect variables in which the manipulability relationships are bidirectional (as in cases of feedback), many, if not most, causal relationships are unidirectional’ (Craver [2007], p. 153, original emphasis).

²⁹ In the literature on causal modeling, causal systems are commonly assumed to be acyclic. This assumption drastically simplifies causal discovery and causal inference (see Pearl [2000]; Spirtes et al. [2000]; see also Leuridan [2009] for a logic of causal discovery that relies on this assumption). Although it is closely connected to this literature, Woodward's framework does not presume such acyclicity (see Woodward [2003], p. 396).

irrelevant to the latter, because it involves changes to some relevant component A (Craver [2007], pp. 148–9). Again, X is not a component but only a background condition. Thus, experiments that are *prima facie* top-down (as the first example) or bottom-up (as in the second example) may fail to establish constitutive relevance because they only involve changes in background conditions. This is the kind of underdetermination Craver addresses. Can this problem be solved? Craver presents four ways in which, in neuroscientific practice, components are distinguished from mere background conditions:

First, sometimes mere background conditions are identified by conjoining interference and stimulation strategies. Intervening to inhibit a background condition B 's ϕ -ing may inhibit S 's ψ -ing, but one cannot stimulate S 's ψ -ing by stimulating B 's ϕ -ing. [...] Second, sometimes background conditions can be ruled out on the basis of activation experiments. Although one can interfere with S 's ψ -ing by interfering with background condition B 's ϕ -ing, at least in many cases, one cannot alter B 's ϕ -ing by manipulating S 's ψ -ing. [...] Third, the effects of interfering with background conditions tend to be nonspecific, that is, they affect many phenomena besides the one under study. [...] Finally, the effects of interventions that change background conditions on the behaviors of mechanisms are often unobvious. (Craver [2007], pp. 157–8)

More generally, Craver argues that bottom-up and top-down experiments are (and should) be conducted in tandem. The inferential challenges faced by one kind of experiment are often resolved by another. We saw that this is part of the reason why he associates constitutive relevance with *mutual manipulability*.

I do not doubt the merits of these strategies to solve Craver's problem of distinguishing between components and mere background conditions. The reason I briefly present them is that they will not in general help to solve Problem #3. Bidirectional causation can be studied by conjoining (what seem to be) interference and stimulation strategies. It is not ruled out by (what seem to be) activation experiments. And it does not tend to be nonspecific or unobvious (at least it is no less specific and/or unobvious than constitutive relevance relations). In the following two sections, I will present two guidelines which, when used in tandem, are fairly good yet fallible means to solve Problem #3.

10 Solving the Problem of Underdetermination: Parthood

An experiment is interlevel if and only if intervention and detection are applied at different levels. X 's ϕ -ing and S 's ψ -ing are at different levels if and only if one is a component in a mechanism for the other (recall that levels are relative to particular mechanisms). This is the case if and only if one is constitutively

relevant for the other. But then the distinction between etiological and inter-level experiments hinges on the notion of constitutive relevance. What ways are there to solve Problem #3 without assuming we already know all constitutive relevance relations?

In my opinion, good evidence of part-whole relations is indispensable (yet not a wonder drug) here, as it provides a partial solution to Problem #3. If it is clear that X is not part of S , then X 's ϕ -ing cannot be a component in a mechanism for S 's ψ -ing. Hence, the bidirectional manipulability relations between the two (if any) must be causal. When is X not part of S ? The following variant of **STI** can help to explicate this:

(**STI**[−]) If there is good evidence that the spatio-temporal region occupied by X is *not* contained in the spatio-temporal region occupied by S , then that is good evidence that X is *not* part of S .³⁰

This solution is only partial. First, it does not help to rule out constitutive relevance in the case of endosymbiosis where our evidence tells us that the *Buchnera* are contained in the aphids. Second, in many cases of bidirectional causation one does not have good evidence that the one causal relatum is not contained in the other. Causation in biology is a messy affair. Thus we should seek for a complementary solution to Problem #3. In the following section, I will discuss one such solution and present two further methodological problems it faces.

11 Synchronicity, the Etiological Nature of Experimental Apparatus, and Causal-Constitutive Propagation

According to Craver and Bechtel, interlevel relations are synchronic whereas on many accounts of causation causes must precede their effects. Thus *prima facie* the following principle could be used as a complementary solution to Problem #3.

(**SYNCHR**) Suppose that an experimental intervention³¹ on X 's ϕ -ing results in a change in S 's ψ -ing. If the change in S 's ψ -ing occurs synchronically with the intervention, then the relation between X 's ϕ -ing and S 's ψ -ing is constitutive. If the change in S 's ψ -ing occurs later in time, the relation is etiological.

In my opinion, **SYNCHR** is a fairly good principle for solving the problem of underdetermination, but it faces two problems. A first, minor problem is that—although rare—there may exist counterexamples to the underlying

³⁰ Just like **STI**, **STI**[−] is only meant as a criterion for parts and wholes that have a spatio-temporal location. Contrary to **STI**, it is a methodological principle.

³¹ **SYNCHR** is about real experimental interventions, not ideal interventions.

idea: cases of synchronic invariant relations. Woodward does not explicitly address the issue,³² but his example of a simple pendulum involves synchronic causal relations. The length l of a simple pendulum is rightly regarded as a cause of that pendulum's period T . By intervening to change l , one changes T (Woodward [2003], p. 197). Yet the change of period does not *follow* the change of length *in time*.

The second problem is methodologically more threatening. It is often difficult to distinguish between synchronic and diachronic effects of interventions in practice. More specifically, the effects of interlevel interventions may *look* diachronic even if they are in fact synchronic. In the rest of this section I will discuss two reasons why this may be the case.

The first reason has to do with the *etiological nature of the apparatus* (both the intervention apparatus and the detection apparatus) in interlevel experiments. In these experiments, an intervention technique is applied to one level of the mechanism; then it is verified by means of some detection technique (measurement apparatus, visualization technique) whether a change occurs at the other level of the mechanism. Both intervention and detection involve causal processes, and thus a time lag.³³ There is a causal process and a time lag between the change in the interlevel intervention variable I and the resulting change in X 's ϕ -ing (resp. S 's ψ -ing). And there is a causal process and a time lag between the change in S 's ψ -ing (resp. X 's ϕ -ing) and the detection of that change. As a result, even if the relation between the change in X 's ϕ -ing (resp. S 's ψ -ing) and the change in S 's ψ -ing (resp. X 's ϕ -ing) is synchronic, the relation between intervention and detection is diachronic. Conversely (and here is the problem of underdetermination again): if the relation between intervention and detection is diachronic, the relation between X 's ϕ -ing (resp. S 's ψ -ing) and S 's ψ -ing (resp. X 's ϕ -ing) may either be diachronic (and hence etiological) or synchronic (and hence most plausibly constitutive).

The second reason is what I would call *causal-constitutive propagation*.³⁴ As an example, suppose that the mechanism for S 's ψ -ing consists of the organized action of X_1 's ϕ_1 -ing, X_2 's ϕ_2 -ing, and X_3 's ϕ_3 -ing, that all and only the following qualitative causal relations hold at the lower level: $X_1\phi_1 \rightarrow X_2\phi_2 \rightarrow X_3\phi_3$, and that the actual values of the respective variables are: $\phi_1 = \phi_1^*$, $\phi_2 = \phi_2^*$ and $\phi_3 = \phi_3^*$. Let us consider now the macro-level effect of

³² Woodward does not address the issue in *Making Things Happen*, ([2003]) and none of the definitions comprising his account refers to temporal priority as a condition for causation. I am not aware of any other source in which he does address the issue.

³³ They would not involve a time lag if they were themselves cases of synchronic causation, but I take it that most if not all intervention/detection techniques are diachronic.

³⁴ Causal-constitutive propagation resembles Kistler's 'delayed' interlevel causation ([2009], pp. 598, 604). But whereas I raise causal-constitutive propagation in a methodological discussion, he raises delayed interlevel causation as a conceptual possibility, based on the notion of 'partial constraint'.

an interlevel intervention I on X_1 's ϕ_1 -ing at three consecutive points in time. First, X_1 's ϕ_1 -ing is intervened upon so that at time t its value is set to $\phi_1 = \phi_1'$ (with $\phi_1' \neq \phi_1^*$). As a result, S 's ψ -ing synchronically (i.e. at time t) obtains some value $\psi = \psi'$ (by CR1). But the change in X_1 's ϕ_1 -ing also has a diachronic etiologial effect on X_2 's ϕ_2 -ing so that at time t' (with $t < t'$) $\phi_2 = g(\phi_1')$, where g represents the causal relation between ϕ_1 and ϕ_2 , and $g(\phi_1') \neq \phi_2^*$. As a result, S 's ψ -ing synchronically (i.e. at time t') obtains some value $\psi = \psi''$ (which may differ from ψ'). Analogously, the change in X_2 's ϕ_2 -ing etiologially affects X_3 's ϕ_3 -ing so that at time t'' (with $t' < t''$) $\phi_3 = g'(g(\phi_1')) \neq \phi_3^*$ and $\psi = \psi'''$ (which again may differ both from ψ' and ψ''). In general, a change in X_1 's ϕ_1 -ing at time t may result in changes in S 's ψ -ing at later times t', t'', \dots even if the relation between X_1 's ϕ_1 -ing and S 's ψ -ing is constitutive.

For these two reasons, even if constitutive relations are synchronic, they may look like diachronic relations in practice. Conversely, relations that seem diachronic in practice may either be genuinely diachronic (and hence etiologial) or synchronic (and hence most plausibly constitutive).

The first problem (cases of synchronic causation) and the second problem (the diachronic appearance of constitutive relations) prompt us to revise **SYNCHR**:

SYNCHR* Suppose that an experimental intervention on X 's ϕ -ing results in a change in S 's ψ -ing. If the resulting change seems to occur synchronically with the intervention, then we have reasons to believe that the relation between X 's ϕ -ing and S 's ψ -ing is constitutive. If the resulting change seems to occur later in time, then we have reasons to believe that the relation is etiologial.

The phrases 'seem to' and 'reasons to believe' capture the methodological pitfalls outlined in this section, without providing grounds for a full-blown skepticism regarding etiologial and/or interlevel experiments. I think that **SYNCHR*** and **STI⁻**, when used in tandem, are fairly good yet fallible guidelines to solve Problem #3. (Note that, in Section 6, my main reason for denying that the *Buchnera* are components in the aphids, was an instance of **SYNCHR***.)

12 Pragmatic Interests and the Boundaries of Mechanisms

In his 'Mechanisms and Natural Kinds' ([2009]), Craver explicitly argues that there is no objective way to define the boundaries of mechanisms, at least not in a strong sense of 'objective' (see also Craver [2007], pp. 141–4).³⁵ Pragmatic

³⁵ Craver's main goal in his 'Mechanisms and Natural Kinds' is to argue against the hopes of the homeostatic property cluster (HPC) view to provide an account of natural kinds that is free of any conventional elements.

concerns, human interests, and perspective co-determine the boundaries of mechanisms because they help determine what is *relevant* and what is not. Could these pragmatic considerations help Craver to solve the problem of underdetermination? Aren't cases of underdetermination just cases where relevance should play a more prominent role?³⁶ I completely endorse the view that pragmatic concerns and perspective play an important role in our mechanistic thinking, but my answer to the above question will be negative. Let me consider two issues to motivate this answer.

A first issue is: relevance *to what*? Craver is very explicit on this point. 'The answer is: relevance to the phenomena that we seek to predict, explain, and control' (Craver [2009], p. 590). In other words, much depends on our choice of explanandum phenomenon and on the way we choose to describe it (Craver [2009], p. 591). This is just a restatement of Glennan's law: every mechanism is a mechanism *of* or *for* a particular phenomenon; and it does not seem to alleviate the problem of underdetermination. On the one hand, in both Sections 6 and 9, I explicitly made sure to stay true to Glennan's law when defining both the aphid's and the *Buchnera*'s relevant activities (and hence the explanandum phenomenon). In the case of the aphid–*Buchnera* symbiosis, it was not the delineation of the explanandum phenomenon that solved the underdetermination, but an application of **SYNCHR***. On the other hand, this emphasis on the explanandum phenomenon does not set mechanistic thinking apart from causal thinking. Nothing is a cause *simpliciter*. Every cause is a cause *of* something else. The same holds for our descriptions of causal phenomena. Whether one variable is a cause of another variable in Woodward's sense, depends both on objective features of this world *and* on how the respective variables are defined.³⁷ Likewise, whether one variable is a direct cause of another variable, depends on which other variables we take into account. How we define our variables and which variables we do or do not take into account, partly depends on our human interests, on the phenomena we want to predict, explain, or control, and what we want to use our knowledge for. In the case of symbiotic relations, for example, issues of manipulation and control explicitly set the research agenda:

As well as being of general entomological interest, the mycetocyte symbioses are potentially of great applied value. Many of the insects that depend on their mycetocyte symbionts for normal growth and reproduction are pests of agricultural or medical importance. They include crop pests [...], timber pests [...], insects of public health

³⁶ The suggestion that cases of underdetermination are cases where relevance should play a more prominent role is due to an anonymous referee. I would like to thank him or her for pressing me on this issue.

³⁷ Woodward defends a realist position with respect to causation ([2003], pp. 118–23), but he also explicitly stresses dependence of our causal conclusions on our choice of representation (*passim*) and the role of pragmatics in causal explanation (pp. 226–33).

importance. [...] To date, the requirement of these various groups for microorganisms has not been exploited in pest management. (Douglas [1998], p. 19)

The second issue is: what kind of relevance? Again, Craver is very explicit. The type of relevance he has in mind is constitutive relevance as defined by the mutual manipulability account (Craver [2009], p. 593, footnote 7). But as we saw in Section 9, the problem of underdetermination seems insoluble when we allow ourselves to use only the resources of the mutual manipulability account. At the minimum, we also need methodological principles based on the notions of parthood (STI^-) and synchronicity ($SYNCHR^*$) to try to solve the problem.

In short, even though I welcome the role of pragmatic considerations, human interests, and perspective in the context of mechanistic thinking, I doubt these considerations set it apart from the context of causal thinking and I deny that they can solve the problem of underdetermination. Relevance, broadly construed, does and should play a prominent role both in the context of interlevel experiments and in the context of etiological experiments.

13 Concluding Remarks

Let me briefly recapitulate. Craver takes the intralevel causal relations in mechanisms as invariant under ideal etiological interventions. At the same time, he offers a mutual manipulability account of the interlevel constitutive relations. This gives rise to two conceptual problems. Problem #1: either constitutive relevance implies causal relevance, or we need a set of principled reasons why ideal interlevel interventions cannot be ideal etiological interventions. Problem #2: if parthood is defined in terms of spatio-temporal inclusion, the mutual manipulability account is prone to many counterexamples and fails to provide a sufficient condition for constitutive relevance. Problem #2 has a methodological counterpart: it may be hard to distinguish between constitutive relevance and bidirectional causation on the basis of experimental evidence (Problem #3). Partial, but fallible solutions to this problem have been formulated, based on the notions of parthood (STI^-) and synchronicity ($SYNCHR^*$).

In my opinion, these problems are of the utmost importance for all those who account for a mechanism's causal relations in terms of Woodward-style invariance (such as Glennan [2002]; Woodward [2002]; Craver [2006], [2007]; Leuridan [2010]). Unless we find a better account of parthood, Problems #1 and #2 prompt us either to reject invariance as an appropriate account of causation in the context of mechanisms, or to revise the concepts of ideal etiological and/or ideal interlevel intervention, or to elaborate the mutual manipulability account such that it includes a synchronicity requirement,

or ... to abandon the strict distinction between constitutive and causal relevance.

I propose to take the last road. If we hold on to the *central manipulationist idea* that causal relationships are distinctive in that they are potentially exploitable for the purposes of manipulation and control, then constitutive relationships are definitely causal. Causal relevance (*sensu latu*) thus comes in two varieties: intralevel etiological relevance (or causal relevance *sensu strictu*) and interlevel relevance. Given that the label ‘constitutive’ suggests the latter cannot be causal, we should drop it.

Surely this will not magic away the above problems. There *is* a difference between interlevel relations and bidirectional etiological relations and we *do* want our experimental practices to reveal such distinctions. But my proposal makes clear that the difference in question is not as strict as Craver and Bechtel claim it to be and that as yet we lack the means to spell it out.

Acknowledgements

I would like to thank the following people (in alphabetical order) for their helpful comments, suggestions, and criticisms: Dries Bonte, Carl Craver, Raoul Gervais, Phyllis Illari, P. D. Magnus, Richard Samuels, Sam Schindler, Erik Weber, Adam White, and Jim Woodward, as well as two anonymous referees of this journal. The author is Postdoctoral Fellow of the Research Foundation – Flanders (FWO). Part of the research for this paper was conducted when he was a Visiting Fellow at the Center for Philosophy of Science (University of Pittsburgh).

Appendix

In this appendix, I show that if I is a Craverian ideal interlevel intervention, satisfying (I1_c)–(I4_c), then it satisfies Woodward’s conditions for being an ideal etiological intervention.³⁸ Since Woodward treats these conditions as both necessary *and* sufficient for being an ideal etiological intervention and since he treats invariance under such interventions as both necessary *and* sufficient for being a causal generalization, this result is even more damaging to Craver and Bechtel’s claims that constitutive relevance is strictly distinct from causal relevance than the argument in Section 5.

Here is Woodward’s definition (subscripts are added to the conditions in the interest of clarity, ‘w’ referring to ‘Woodward’; ϕ and ψ are substituted for X

³⁸ In line with my remark in Section 3, I will sidestep Woodward’s distinction between interventions and intervention variables since it does not affect the argument.

and Y ; and in line with footnote 15, I will assume it makes sense to talk about causal intermediates between I and ϕ and ψ):

I is an intervention variable for ϕ with respect to ψ if and only if I meets the following conditions:

(I1_w) I causes ϕ .

(I2_w) I acts as a switch for all the other variables that cause ϕ . That is, certain values of I are such that when I attains those values, ϕ ceases to depend on the values of other variables that cause ϕ and instead depends only on the value taken by I .

(I3_w) Any directed path from I to ψ goes through ϕ . That is, I does not directly cause ψ and is not a cause of any causes of ψ that are distinct from ϕ except, of course, for those causes of ψ , if any, that are built into the I - ϕ - ψ connection itself; that is, except for (a) any causes of ψ that are effects of ϕ (i.e. variables that are causally between ϕ and Y) and (b) any causes of ψ that are between I and ϕ and have no effect on ψ independently of ϕ .

(I4_w) I is (statistically) independent of any variable Z that causes ψ and that is on a directed path that does not go through ϕ (adapted from Woodward [2003], p. 98).

Here is the proof. Suppose that I is an ideal interlevel intervention. Then, first, it satisfies (I1_w) by the general characterization of interlevel interventions. Second, it also satisfies (I2_w) by (I4_c). Third, I satisfies a condition which is slightly stronger than (I3_w). By (I1_c), it does not directly cause ψ . And by (I2_c) it does not change the value of any ϕ^* that changes the value of ψ , except via ϕ . Therefore, it is not a cause of any causes of ψ that are distinct from ϕ and that are not built into the I - ϕ - ψ connection. Fourth, I is (statistically) independent of any variable Z that causes ψ and that is on a directed path that does not go through ϕ , as required by (I4_w). The reason is that any such Z is causally independent of I (since by (I3_w) any directed path from I to ψ goes through ϕ); hence I is statistically independent from Z by (I3_c).

In short, if I is an ideal interlevel intervention, it satisfies the (necessary) conditions (I1_c)–(I4_c). Therefore, it satisfies the (necessary and sufficient) conditions (I1_w)–(I4_w). Thus it is a Woodwardian ideal etiological intervention.

*Centre for Logic and Philosophy of Science
Department of Philosophy and Moral Sciences
Ghent University
Ghent
Belgium
Bert.Leuridan@UGent.be*

References

Bechtel, W. [2008]: *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*, New York: Taylor & Francis Group LLC.

- Bechtel, W. and Abrahamsen, A. [2005]: 'Explanation: A Mechanist Alternative', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**, pp. 421–41.
- Cartwright, N. [2002]: 'Against Modularity, the Causal Markov Condition, and any Link between the Two: Comments on Hausman and Woodward', *British Journal for the Philosophy of Science*, **53**, pp. 411–53.
- Craver, C. F. [2002]: 'Interlevel Experiments and Multilevel Mechanisms in the Neuroscience of Memory', *Philosophy of Science*, **69**, pp. S83–97.
- Craver, C. F. [2006]: 'When Mechanistic Models Explain', *Synthese*, **153**, pp. 355–76.
- Craver, C. F. [2007]: *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Oxford University Press.
- Craver, C. F. [2009]: 'Mechanisms and Natural Kinds', *Philosophical Psychology*, **22**, pp. 575–94.
- Craver, C. F. and Bechtel, W. [2007]: 'Top-down Causation without Top-down Causes', *Biology and Philosophy*, **22**, pp. 547–63.
- Douglas, A. E. [1993]: 'The Nutritional Quality of Phloem Sap Utilized by Natural Aphid Populations', *Ecological Entomology*, **18**, pp. 31–8.
- Douglas, A. E. [1998]: 'Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria *Buchnera*', *Annual Review of Entomology*, **43**, pp. 17–37.
- Glennan, S. S. [1996]: 'Mechanisms and the Nature of Causation', *Erkenntnis*, **44**, pp. 49–71.
- Glennan, S. S. [2002]: 'Rethinking Mechanistic Explanation', *Philosophy of Science*, **69**, pp. S342–53.
- Goodman, N. [1977]: *The Structure of Appearance*, 3rd edition., Dordrecht: Reidel.
- Hausman, D. and Woodward, J. [1999]: 'Independence, Invariance and the Causal Markov Condition', *British Journal for the Philosophy of Science*, **50**, pp. 521–83.
- Hausman, D. and Woodward, J. [2004]: 'Modularity and the Causal Markov Condition: A Restatement', *British Journal for the Philosophy of Science*, **55**, pp. 147–61.
- Ishikawa, H. [2003]: 'Insect Symbiosis: An Introduction', in K. Bourtzis and T. A. Miller (eds), *Insect Symbiosis*, Boca Raton: CRC Press, pp. 1–21.
- Kauffman, S. A. [1971]: 'Articulation of Parts Explanation in Biology and the Rational Search for Them', in R. C. Buck and R. S. Cohen (eds), *PSA 1970*, Dordrecht: Reidel.
- Kistler, M. [2009]: 'Mechanisms and Downward Causation', *Philosophical Psychology*, **5**, pp. 595–609.
- Koslicki, K. [2008]: *The Structure of Objects*, Oxford: Oxford University Press.
- Leuridan, B. [2009]: 'Causal Discovery and the Problem of Ignorance: An Adaptive Logic Approach', *Journal of Applied Logic*, **7**, pp. 188–205.
- Leuridan, B. [2010]: 'Can Mechanisms Really Replace Laws of Nature?', *Philosophy of Science*, **77**, pp. 317–40.
- Lombardo, M. P. [2008]: 'Access to Mutualistic Endosymbiotic Microbes: An Underappreciated Benefit of Group Living', *Behavioral Ecology and Sociobiology*, **62**, pp. 479–97.

- Machamer, P., Darden, L. and Craver, C. F. [2000]: 'Thinking about Mechanisms', *Philosophy of Science*, **67**, pp. 1–25.
- Martin, E. and Hine, R. [2008]: *A Dictionary of Biology*. Oxford University Press; *Oxford Reference Online* <www.oxfordreference.com>.
- Meek, C. [1995]: 'Strong Completeness and Faithfulness in Bayesian Networks', in S. Hanks and P. Besnard (eds), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, San Francisco: Morgan Kaufmann, pp. 411–8.
- Mitchell, S. D. [2003]: *Biological Complexity and Integrative Pluralism*, Cambridge: Cambridge University Press.
- Mitchell, S. D. [2009]: *Unsimple Truths: Science, Complexity and Policy*, Chicago: University of Chicago Press.
- Pâslaru, V. [2009]: 'Ecological Explanation between Manipulation and Mechanism Description', *Philosophy of Science*, **76**, pp. 821–37.
- Pearl, J. [2000]: *Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.
- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. and Ishikawa, H. [2000]: 'Genome Sequence of the Endocellular Bacterial Symbiont of Aphids *Buchnera* Sp. APS', *Nature*, **407**, pp. 81–6.
- Spirites, R., Glymour, C. and Scheines, R. [2000]: *Causation, Prediction, and Search*, Cambridge, Mass: MIT Press.
- Steel, D. [2006]: 'Comment On Hausman and Woodward on the Causal Markov Condition', *British Journal for the Philosophy of Science*, **57**, pp. 219–31.
- Varzi, A. [2009]: 'Mereology', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*; <<http://plato.stanford.edu>>.
- Woodward, J. [2002]: 'What is a Mechanism? A Counterfactual Account', *Philosophy of Science*, **69**, pp. S366–77.
- Woodward, J. [2003]: *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.
- Woodward, J. and Hitchcock, C. [2003]: 'Explanatory Generalizations, Part I: A Counterfactual Account', *Noûs*, **37**, pp. 1–24.