The Causal Situationist Account of Constitutive Relevance*

Emily Prychitko**

Abstract An epistemic account of constitutive relevance lists the criteria by which scientists can identify the components of mechanisms in empirical practice. Three prominent claims from Craver (2007) form a promising basis for an account. First, constitutive relevance is established by means of interlevel experiments. Second, interlevel experiments are executions of interventions. Third, there is no interlevel causation between a mechanism and its components. Currently, no account on offer respects all three claims. I offer my causal situationist account of constitutive relevance that respects the claims. By situating a part of a mechanism on the causal chain between the mechanism's input and output, components can be identified with interventions, without the interventions suggesting interlevel causation. The causal situationist account is the only account on offer so far that clearly fits within Craver's (2007) framework.

1 Introduction

Through a series of experiments, scientists have discovered that the middle temporal visual area (MT) is a component of the mechanism for motion perception. MT cells are active when motion is being perceived (Britten et al. 1992), inhibition of MT cells impairs one's ability to perceive motion (Newsome and Paré 1988), and direct stimulation of MT cells makes one perceive a particular direction of motion (Salzman et al. 1990). Because MT cells are a physical part of the neural mechanism for motion perception and their activity is relevant to the perceiving of motion, any account of constitutive relevance should yield that MT cells are a component of the mechanism for motion perception.

An epistemic account of constitutive relevance lists the criteria by which scientists can determine, in empirical practice, whether something is a component of a mechanism. Such an account is essential for distinguishing the components of the mechanism from mere parts or causes of the mechanism's behavior.

There are three claims that provide a promising basis for an account of constitutive relevance, as originally put forth by Craver (2007). First, in practice, constitutive relevance between a mechanism and its components is discovered by way of interlevel experiments (Craver 2007; Craver and Bechtel 2007; Kaplan 2012). Second, like experiments that establish causal relevance, interlevel experiments involve Woodwardian interventions (Craver 2007; Craver and Bechtel 2007; Baumgartner and Gebharter 2015; Krickel 2018). Third, constitutive relevance is, nevertheless, distinct from, rather than a type of, causal relevance (Craver 2007; Romero 2015; Baumgartner and Gebharter 2015; Krickel 2018).

Though there are several accounts of constitutive relevance on offer, none of them respect all three claims. Some authors even suggest that it is impossible for any account to

* March 2019. Forthcoming in *Synthese*, https://doi.org/10.1007/s11229-019-02170-4
** Philosophy-Neuroscience-Psychology, Washington University in St. Louis; eprychi@wustl.edu

respect all three (Baumgartner and Casini 2017). I will propose my 'causal situationist' account of constitutive relevance which satisfies the three claims straightforwardly.

2 Constraints for an Account of Constitutive Relevance

The mechanism for a phenomenon consists in the entities and activities that instantiate the phenomenon when they are causally and spatiotemporally organized in the right way[1] (Craver 2007; see Glennan 2002 and Bechtel and Abrahamsen 2005 for other definitions). Where S is the mechanism and ψ-ing is its behavior of interest, S's ψ-ing is the occurrence of the phenomenon. S's ψ-ing is constituted by the organized activities of S's components, Xs' Φ-ings. The relationship between the mechanism and its components is both spatiotemporal and active. That is, physical parts of S are components only if their behaviors are constitutively relevant to the ψ-ing of S.

We can think of S's ψ-ing as consisting in the causal chain between S's input or startup conditions (S_{IN}) and S's output or termination conditions (S_{OUT}). S_{IN} is that which causes S to ψ, that is, causes the Xs to Φ. This then causes an output to appear. The basic structure of S's ψ-ing is captured in Fig. 1.

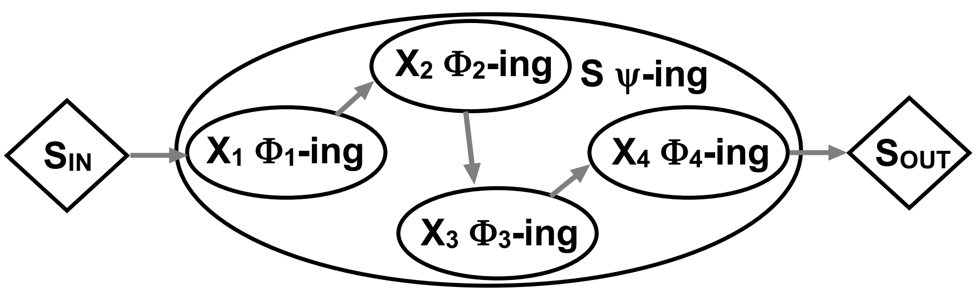

Fig. 1 A mechanism. S's ψ-ing is the explanandum phenomenon, and consists in Xs' Φ-ings. Xs' Φ-ings constitute a causal chain from S_{IN} to S_{OUT}. Adapted from Craver (2007, p. 7).

Presenting a subject with a moving stimulus (S_{IN}), for instance, causes her to perceive motion. The perceiving of motion (S's ψ-ing) consists in the movement of light through the eyeball, the transduction of energy through the retina, the traveling of information through the optic nerve and thalamus, the building of representations in the visual areas (Xs' Φ-ings); these active entities are causally related to each other. Once the subject has perceived motion, this causes her to, for instance, say she saw something move (S_{OUT}).

This is, of course, a simplistic characterization of mechanisms. Any mechanism will likely have numerous inputs and outputs and different particular manners of ψ-ing. The spatiotemporal boundaries of the mechanism may not be clear. Some components may stand in causal loops. A set of known and perhaps unknown background conditions,

[1] Assuming the relevant background conditions are present. One's heart must be beating, for example, for one's mechanism for motion perception to operate at all.

unrepresented in the diagram, are required for the mechanism to work in the first place. Whether something counts as a background condition or as an input to the mechanism may be somewhat arbitrary. Nevertheless, the simplistic characterization will do for our purposes, since it highlights the core of many mechanisms (Craver 2007).[2]

The goal of an epistemic account of constitutive relevance is to tell us how to identify the components of a mechanism in practice. It lists the criteria by which one can show, empirically, that something is a component of a mechanism. Multiple accounts may provide sufficient conditions for the empirical establishment of constitutive relevance. Even if there is one correct set of necessary and sufficient conditions for whether something is, in fact, a component of a mechanism, there may be several actual or merely possible practices that scientists can use to discover that something is a component. I will offer one epistemic account, but recognize that other accounts may highlight different empirical practices that are capable of identifying components just as well.

For the purposes of this paper, however, I will focus on three claims prominent in Craver (2007) and treat them as constraints on an account of constitutive relevance, with the hope of finding an account that meets them. There are four reasons for this project. First, as I will argue, the constraints form at least one promising basis for an account. Second, as I will show, there is currently no account that meets the constraints. Third, it is worth determining if it is even possible for a good account to respect the constraints; if not, that may point to a problem with Craver's framework. Fourth, proponents of Craver's work may have an interest in discovering an account of constitutive relevance that fits in his framework.

The first claim from Craver is that scientists discover the components of mechanisms by executing multiple interlevel experiments (Craver 2007; Craver and Bechtel 2007; Kaplan 2012). Craver (2007) spells out the experiments in detail. In top-down experiments, scientists manipulate the mechanism as a whole to see if this thereby creates a change in the putative component, which must be a physical part of the mechanism. For example, Britten et al. (1992) showed monkeys an array of dots, where a proportion of the dots moved in the same direction while the rest moved in random directions. Meanwhile, the activity of the monkeys' MT cells was recorded with an electrode. After being shown the array, the monkeys had to indicate in which direction the dots had moved by looking towards the direction of motion. Britten et al. found that the firing rate of the monkeys' MT cells accounted for the accuracy of the monkeys' judgments. The higher the firing rate, the more accurate the monkeys' judgments tended to be.

In bottom-up experiments, scientists manipulate the putative component to see if this creates a change in the mechanism as a whole. For example, Nichols and Newsome (2002) did an experiment with a similar design as Britten et al., but on some trials they stimulated the MT cells while the array was being viewed by the monkeys. They found that stimulation of the MT cells significantly influenced the monkeys' judgments of the direction in which the dots had moved. Another type of bottom-up experiment involves inhibiting, rather than activating, the putative component. Newsome and Paré (1988) also ran a similar experiment, but they lesioned the monkeys' MT cells unilaterally. They found that this increased the monkeys' threshold for motion perception in the lesioned hemifield. A greater

[2] One noteworthy exception is an amplifier whose transistors operate in parallel, and thus do not comprise a single causal chain, but are nevertheless components (Wimsatt 2007, pp. 281-287).

proportion of the dots had to be moving in the same direction than before the lesions for the monkeys to judge the direction of motion accurately.

Interlevel experiments gauge the relationship between a mechanism and some of its parts. They screen out the parts that are related to the mechanism merely spatiotemporally. Melanin in the iris, for example, is a mere part of the mechanism for motion perception. Top-down and bottom-up experiments in which the behavior of the putative component accompanies the behavior of the mechanism suffice jointly for constitutive relevance.

Top-down experiments alone are not sufficient. Suppose a top-down experiment shows that the ventral striatum is active when an array of moving dots is presented to a subject. It might be that the ventral striatum is active only because the subject expects to get rewarded for judging the direction of motion accurately. Were that the case, the ventral striatum might be a component of the mechanism for reward learning, but not for motion perception per se.

Bottom-up experiments alone are also not sufficient. Lesioning the motor cortex of a subject might decrease the accuracy of his judgments in a motion perception task. That may simply be because the activity of the motor cortex is necessary for the subject to perform his response, instead of for him to perceive motion. Top-down and bottom-up experiments together ensure that the same part is behaving in a constitutively relevant way to the mechanism whenever the part changes or the mechanism as a whole changes.

The structure of interlevel experiments is similar to experiments that establish causal relevance (Craver 2007; Craver and Bechtel 2007). Both involve manipulations of some variable with respect to another in order to determine whether an explanatory relationship exists between the variables. Because of this similarity, it is promising to build an account of constitutive relevance from an account of causal relevance.

Woodward's popular (2003) interventionist account of causation, for example, enables us to spell out causal experiments in terms of 'ideal interventions'. Some variable, I, counts as an intervening variable for a variable, A, with respect to another variable, B, if and only if

(I1) I is a cause of A.
(I2) I acts as a switch for all other causes of A.
(I3) Any causally directed path from I to B goes through A.
(I4) I is statistically independent of any variable, C, that causes B and that is on a directed path which does not go through A (adapted from Woodward 2003, p. 98).

I's taking on a particular value is an ideal intervention on A with respect to B, then, if and only if I's taking on that value causes A to take on a certain value, and I counts as an intervening variable for A with respect to B (ibid.). For Woodward, an ideal intervention on A with respect to B, where all variables outside of this causal path are held constant, is necessary and sufficient for A to be a direct cause of B (for more detail, see Woodward 2003, p. 59).

Because we are interested in discovering constitutive relevance in practice, we will say that if scientists closely approximate an ideal intervention on A with respect to B, they can conclude that A is causally relevant to B. This is because they will have shown that the change in A, rather than changes in confounding variables, is likely what produced the change they later observed in B.

The manipulations in successful interlevel experiments can be formulated as executions of (or at least close approximations of) ideal interventions. Craver (2007) originally conceived of successful top-down experiments as executions of ideal interventions on S's ψ-ing with respect to X's Φ-ing, and successful bottom-up experiments as ideal interventions on X's Φ-ing with respect to S's ψ-ing. The idea was that these ideal interventions require scientists to hold potential confounding variables constant, so as to observe the relationship solely between S's ψ-ing and putative components.

Since Craver, other formulations of interlevel experiments in interventionist terms have come about. Some of these spell out ideal interventions as taking place between variables other than S's ψ-ing and X's Φ-ing. Some draw on Woodward's (2015) extended version of interventionism, which posits what I will call 'ideal* interventions'. Ideal* interventions are the same as ideal interventions, except that instead of the intervening variable having to meet (I3) and (I4), in addition to (I1) and (I2), it must meet

> (I3*) Any causally directed path from I to B goes through A, or through some variable that is related to A by supervenience.
>
> (I4*) I is statistically independent of any variable, C, that causes B, that is on a directed path which does not go through A, and that is not related by supervenience to A (adapted from Woodward 2015, p. 333).

Ideal* interventions suffice for causal relevance. Unlike ideal interventions, they do not require scientists to hold variables that stand in supervenience relations to A constant. Because of the looser constraints, there can be more ideal* than ideal interventions, so the extended version of interventionism admits more causal relations than the original version. As we will see, some authors appeal to ideal* interventions in their accounts of constitutive relevance.

The second claim is, then, that interlevel experiments are executions of interventions (on variables that must be specified). The interventions may or may not be ideal/ideal*, depending on how successful the experiments are. Any construal of interlevel experiments in terms of interventions is prima facie attractive, since it accounts for the similarity between the empirical discovery of constitutive and causal relevance. An account of constitutive relevance that interprets interlevel experiments in non-interventionist terms would need to explain away the apparent similarity of the manipulations that scientists execute in causal and interlevel experiments, and argue that the empirical practices themselves are significantly different. An account that requires ideal/ideal* interventions also automatically demands the rigor of holding potential confounding variables constant. Without holding potential confounds constant, the interventions would not count as ideal/ideal*, so the account would not be satisfied. This is an easy way for an account to prevent certain non-components from counting mistakenly as constitutively relevant to a mechanism.

The third claim is that even though constitutive and causal relevance are both discovered by means of ideal/ideal* interventions, constitutive relevance is not causal. There is no interlevel causation between a mechanism and its components (Craver and Bechtel 2007; Romero 2015; Craver 2007; Baumgartner and Gebharter 2015). Causes are typically seen as distinct spatiotemporal events from their effects. But the changes in a mechanism and its components overlap both spatially and temporally, given that the mechanism consists in the components. The perceiving of motion and the activity of MT cells are not causally related, since the perceiving of motion partially consists, spatiotemporally and actively, in the activity of MT cells. Neither event is independent from the other in the way that is required for events to be causally related.

These, then, are the three claims from Craver (2007) that I will treat as constraints on an account of constitutive relevance:

(1) Components are identified by means of interlevel experiments.
(2) Interlevel experiments are executions of interventions.
(3) There is no interlevel causation.

To fully establish the truth of these claims, each would need more support. While I believe there is a strong case to be made in favor of each claim (see, e.g., Craver 2007; Craver and Bechtel 2007), my present concern is to find an account of constitutive relevance that respects all three. As such, I will treat the claims as given.

In addition to each claim having independent motivation, using this set of claims as constraints on an account of constitutive relevance is promising. An account that respects the three constraints can potentially recognize and make sense of existing empirical practices that actually establish constitutive relevance, as well as provide guidance about how one could identify components. While there may be innumerable practices that *could* establish constitutive relevance, formulating an account on, in part, actual empirical practice makes it more likely that the account will be both feasible and fruitful empirically. Explaining how the interventions executed in interlevel experiments serve to identify components, while maintaining that the interventions do not show interlevel causation, can yield an account with some degree of both descriptive and theoretical adequacy.

3 Accounts of Constitutive Relevance

I will argue that none of the epistemic accounts of constitutive relevance that are currently on offer satisfy all three constraints, even though some are presented as doing so.

Craver's (2007) mutual manipulability account of constitutive relevance requires an ideal intervention on S's ψ-ing with respect to X's Φ-ing (in top-down experiments) and on X's Φ-ing with respect to S's ψ-ing (in bottom-up experiments), where X is a spatiotemporal part of S, for X's Φ-ing to count as a component of S's ψ-ing. As many have argued, any intervention on S's ψ-ing with respect to X's Φ-ing cannot be ideal (Baumgartner and Gebharter 2015; Romero 2015; Harinen 2018; Krickel 2018). Ideal interventions require, per criterion (I3), that scientists alter the second variable of interest only through initially intervening on the first variable of interest. An intervention on both X's Φ-ing and S's ψ-ing at the same time is not ideal, since the change to X's Φ-ing would not have been caused by the change in S's ψ-ing. Because X's Φ-ing is constitutively relevant to S's ψ-ing, any intervention on S's ψ-ing, when X's Φ-ing changes, is also a direct, simultaneous intervention on X's Φ-ing. Since this intervention cannot be ideal, the criteria of Craver's account can never be met by the components of mechanisms.[3]

While Craver's account was put forth as an account that respects all three constraints, it in fact forces us to reject at least one. We must deny that the criteria of his account are met in practice. We might then reject (1), holding that interlevel experiments actually do not enable scientists to discover components. Alternatively, we might deny (2) and maintain that while interlevel experiments do allow for the identification of components, they do not do so by means of interventions. We would have to explicate interlevel experiments in other terms.

Baumgartner and Gebharter (2015) require an ideal* intervention on S's ψ-ing with respect to X's Φ-ing (in top-down experiments) and on X's Φ-ing with respect to S's ψ-ing (in bottom-up experiments), where X is a spatiotemporal part of S, in addition to evidence that every variable that causes a change in S's ψ-ing simultaneously causes a change in at least one X's Φ-ing in a set of putative components. Notice that ideal* interventions between S's ψ-ing and X's Φ-ing are indeed possible, since criteria (I3*) and (I4*) allow the intervening variable to change both variables directly. However, these ideal* interventions would suggest that S's ψ-ing and X's Φ-ing are causally related, since ideal/ideal* interventions suffice for causal relevance (Woodward 2003; Baumgartner and Gebharter 2015; Harinen 2018). Baumgartner and Gebharter, to keep from accepting interlevel causation, introduce a principle of causation that says that for two variables to be causally related, it must be possible to induce a change in the first before this thereby creates a change in the second. This is not possible for S's ψ-ing and X's Φ-ing, so interlevel causation is blocked. Baumgartner and Gebharter, therefore, respect (3).

They fail to fulfill (1), however, since they maintain that interlevel experiments are not sufficient for establishing constitutive relevance. Interlevel experiments do not establish that *every* cause of S's ψ-ing is also a cause of some X's Φ-ing. Scientists, as a matter of fact, have rarely if ever found *all* of the causes of any S's ψ-ing, but they have nevertheless identified components of mechanisms. They have done so by executing interlevel experiments. Exactly how many interlevel experiments must be performed and how many different inputs to S's ψ-ing they must make use of in order to identify components are worthwhile questions. But if we want to maintain that scientists have in fact identified components, we cannot hold that establishing constitutive relevance requires showing that *every* cause of S's ψ-ing simultaneously changes some X's Φ-ing.

Krickel (2018) requires an ideal or ideal* intervention on a temporal part of S's ψ-ing with respect to X's Φ-ing (in top-down experiments) and on X's Φ-ing with respect to a temporal part of S's ψ-ing (in bottom-up experiments), where X is a spatiotemporal part of S. Though Krickel presents the account as maintaining (3), it actually does not. The ideal/ideal* interventions are performed on the same physical entity, just at different times; they are on the whole (at a time) with respect to the part, and on the part with respect to the whole (at a time). But the part is part of the whole at all times. Granting causal relations between the whole (at any time) and the part is, then, granting interlevel causation. It is granting that the whole and its parts causally interact. What this series of ideal/ideal*

[3] For in-depth explanations of why interventions on S's ψ-ing with respect to X's Φ-ing cannot be ideal, see Romero (2015) and Baumgartner and Gebharter (2015, pp. 737-745).

interventions gives us is a web of interlevel causal relations between S's ψ-ing and Xs' Φ-ings. But we do not want to admit any causal relations between them. The whole set of Xs' Φ-ings is (at least token) identical to S's ψ-ing, rather than causally related to S's ψ-ing at all.

Harinen (2018) demands an ideal* intervention on S_{IN} with respect to X's Φ-ing (in top-down experiments) and on X's Φ-ing with respect to S_{OUT} (in bottom-up experiments), where X is a part of S. He concludes that there is interlevel causation, rejecting (3), since he thinks of S_{IN} and S_{OUT} as supervening over entities and activities that lie on the same mechanistic level as Xs' Φ-ings.

Harbecke (2015), Gebharter (2016), and Baumgartner and Casini (2017) give accounts that do not appeal to ideal/ideal* interventions at all, so they effectively deny (2). They also require methods other than interlevel experiments for establishing constitutive relevance, so they might deny (1) as well.

The current accounts of constitutive relevance, then, fail to fulfill the three constraints. In fact, the difficulty of creating a satisfying account that fits within Craver's framework has led Baumgartner and Casini (2017) to say the following.

> Since constitution is a noncausal form of dependence – as commonly assumed in mechanistic theorizing – one cannot simply tweak a successful account of causation to obtain a successful account of constitution. Rather, constitution must be defined within a theoretical framework that reflects its distinctly noncausal nature. Furthermore, the inference to constitution can neither in theory nor in practice proceed along the lines of the inference to causation. The main reason is that, while there exist ideal experimental designs allowing for the generation of unconfounded data that conclusively establish the existence of causal relations, no such experimental designs exist for the inference to constitution . . . Hence, the inference to constitution is inherently underdetermined by experimental evidence. (pp. 215-216)

The whole endeavor, according to this view, of trying to create an account of constitutive relevance that respects the three constraints is futile. Constitutive relevance simply cannot be established by interventions executed in interlevel experiments, given that it is not a causal relationship.

Contra Baumgartner and Casini, I will now offer my causal situationist account of constitutive relevance, which respects the three constraints straightforwardly. It allows for the empirical identification of components by way of ideal/ideal* interventions, without entailing interlevel causation.

4 The Causal Situationist Account

If X's Φ-ing is an actual component of S's ψ-ing, interlevel experiments involve changes to at least four variables overall. In top-down experiments, scientists make S ψ by changing S_{IN}, creating differences in X's Φ-ing and S_{OUT}. In bottom-up experiments, scientists might make X Φ, creating differences in S's ψ-ing and S_{OUT}.

Imagine a top-down experiment as involving an intervention on S_{IN} with respect to X's Φ-ing (where X is a part of S), instead of, say, an intervention on S's ψ-ing with respect to X's Φ-ing. S_{IN} and X's Φ-ing are merely causally related, since the presence of S_{IN} triggers the causal procession of the components' activities, one of which is this X's Φ-ing. S_{IN} and X's Φ-ing are not constitutively related, since X's Φ-ing is a component of S's ψ-ing, not S_{IN}. S_{IN} is not a part of S's ψ-ing, because S_{IN} is that which causes S to ψ. The relationship between S_{IN} and X's Φ-ing is a perfectly normal causal relationship. This intervention, then, can be ideal.

The difference between conceiving of a top-down experiment as involving an intervention on S's ψ-ing with respect to X's Φ-ing (a la Craver 2007 and Baumgartner and Gebharter 2015), and an intervention on S_{IN} with respect to X's Φ-ing (a la Harinen 2018 and myself) is captured in Fig. 2.

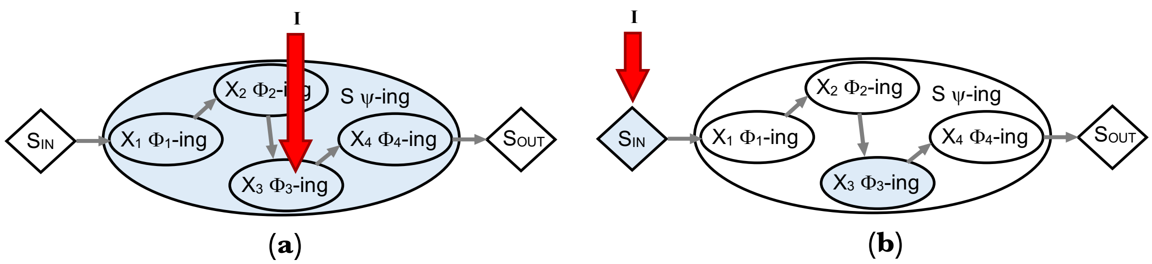

Fig. 2 In (**a**), an intervention on S's ψ-ing with respect to X's Φ-ing, the intervening variable, I, causes a change in both S's ψ-ing and X_3's Φ_3-ing directly. In (**b**), an intervention on S_{IN} with respect to X's Φ-ing, I causes a change in only S_{IN} directly, which then causes a change in X_3's Φ_3-ing.

Here, the arrow points to where the intervening variable (I) acts causally, and the shaded areas represent the two variables between which the intervention is said to occur. In Fig. 2a, the intervening variable acts on both S's ψ-ing and X_3's Φ_3-ing directly, so the intervention cannot be ideal. In Fig. 2b, the intervening variable acts only on S_{IN} directly, and the change in S_{IN} then causes a change in X_3's Φ_3-ing, so the intervention can be ideal.[4]

Imagine, as well, a bottom-up experiment as involving an intervention on X's Φ-ing with respect to S_{OUT}, rather than on X's Φ-ing with respect to S's ψ-ing. Because S_{OUT} lies outside of the boundaries of S's ψ-ing, X's Φ-ing and S_{OUT} are solely causally related. However, in this case, it is not clear that the intervention can be ideal. For the intervention on X's Φ-ing with respect to S_{OUT} to be ideal, any causal path from the intervening variable to S_{OUT} must go through X's Φ-ing causally, given criterion (I3). The intervening variable causes a change in both S's ψ-ing and X's Φ-ing, given their constitutive relationship. But the causal path from S's ψ-ing to S_{OUT} does not include X's Φ-ing. That is, it is not the case that the intervening variable causes a change in S's ψ-ing, which then causes a change in X's Φ-ing, which then causes a change in S_{OUT}. Therefore, the intervention on X's Φ-ing

[4] Diagrams of the interventions performed in interlevel experiments that are similar to Fig. 2b can be found in Bechtel (in press).

with respect to S_{OUT} cannot be ideal. The constitutive relationship between S's ψ-ing and X's Φ-ing violates (I3) in interventions on X's Φ-ing with respect to S_{OUT}.

For this reason, we must instead require an ideal* intervention on X's Φ-ing with respect to S_{OUT}. This intervention is possible, since (I3*) allows the intervening variable to cause S's ψ-ing, which is related to X's Φ-ing by supervenience, which then causes S_{OUT}. (I3*) doesn't require X's Φ-ing to be another joint on the causal path from the intervening variable, to S's ψ-ing, to S_{OUT}.

Conceiving of the interventions executed in interlevel experiments as occurring on S_{IN} with respect to X's Φ-ing and on X's Φ-ing with respect to S_{OUT} makes sense of constitutive relevance. Demonstrating that a part of S is caused by S_{IN} and causes S_{OUT} establishes that it is a component of S's ψ-ing. S's ψ-ing consists in the entities and activities on the causal chain between S_{IN} and S_{OUT}. Showing that something lies on this causal chain would establish that the part is constitutively relevant to S's ψ-ing.

Furthermore, this conception of the interventions does not entail interlevel causation. It would only entail that S_{IN} is causally relevant to X's Φ-ing and that X's Φ-ing is causally relevant to S_{OUT}. Any component of S's ψ-ing stands in these causal relationships with S_{IN} and S_{OUT}. Neither ideal/ideal* intervention would establish that X's Φ-ing is causally relevant to S's ψ-ing.

Here, then, are the criteria of my causal situationist account. X's Φ-ing is a component of S's ψ-ing if

(CS1) X is a part of S.
(CS2) An ideal intervention on S_{IN} changes X's Φ-ing.
(CS3) An ideal* intervention on X's Φ-ing changes S_{OUT}.

The idea is that to show that X's Φ-ing is a component of S's ψ-ing, one must situate X's Φ-ing on the causal chain between S_{IN} and S_{OUT}. That is what the interventions are for.

These criteria are very similar to those of Harinen's (2018) account, but our accounts differ in important ways. Harinen appeals to ideal* rather than ideal interventions for the same reasons that he accepts interlevel causation. Namely, he takes S_{IN} and S_{OUT} to supervene on entities and activities that are on the same mechanistic level as Xs' Φ-ings. Whether S_{IN} and S_{OUT} supervene over entities and activities *other than* Xs' Φ-ings is not made clear on his account, but either way, this position is problematic.

First, it is problematic to think of S_{IN} and S_{OUT} as supervening over any of Xs' Φ-ing, for they do not. They are not parts of S's ψ-ing. Instead, they are defined by their causal relations to S's ψ-ing: S_{IN} is that which causes S to ψ, and S_{OUT} is the causal effect of S's ψ-ing. It seems generally to be the case that an input to and output from a system are not constitutive of the system, but are causes and effects of the system's behavior. This being so, considering them parts of the system would itself entail part-whole causation, which we are trying to avoid. Keeping S_{IN} and S_{OUT} outside of the boundaries of S's ψ-ings allows us to say that the ideal/ideal* interventions of the causal situationist account do not suggest interlevel causation.

If, instead, Harinen takes S_{IN} and S_{OUT} to supervene on only *non-components* of S's ψ-ing, that problem can be avoided. But it would still be incorrect to infer interlevel causation.

One thing lies on a higher mechanistic level than another thing if and only if the latter is a component of the former (Povich and Craver 2017). Even if S_{IN} and S_{OUT} supervene on non-components, any X's Φ-ing is not a component of S_{IN} nor S_{OUT}. So, S_{IN} and S_{OUT} do not lie on a higher mechanistic level than any X's Φ-ing. Furthermore, two things are on the same mechanistic level only if they are components of the same mechanism (ibid.). The components and non-components of S's ψ-ing thus do not lie on the same mechanistic level. Even if S_{IN} and S_{OUT} were on a higher mechanistic level than non-components of S's ψ-ing, given that the non-components and components of S's ψ-ing are not on the same mechanistic level, S_{IN} and S_{OUT} still would not lie on a higher level than any X's Φ-ing. For these reasons, there is no interlevel relationship between any X's Φ-ing and S_{IN} or S_{OUT}. Ideal/ideal* interventions on these variables, then, do not establish *interlevel* causation.

The causal situationist account, unlike Harinen's account, respects the three constraints. First, it maintains that components are identified through interlevel experiments. We have not changed our conception of how the interlevel experiments themselves are conducted empirically. Second, it maintains that interlevel experiments are executions of interventions. Third, it can maintain that there is no interlevel causation, since the ideal/ideal* interventions on S_{IN} with respect to X's Φ-ing and on X's Φ-ing with respect to S_{OUT} do not indicate that there is interlevel causation.

Consider again the perceiving of motion, S's ψ-ing, and the activity of MT cells, X's Φ-ing. What counts as S_{IN} and S_{OUT}, and the spatiotemporal boundaries of S's ψ-ing, is somewhat a matter of choice. Intuitively, though, we might think of S's ψ-ing as including the head spatiotemporally, S_{IN} as including the presentation of a moving stimulus, and S_{OUT} as including a behavioral response that a subject has perceived motion.

By the causal situationist account, then, we would say that top-down experiments are interventions on the subject's being shown a moving stimulus with respect to the activity of the subject's MT cells. In the Britten et al. (1992) study in which monkeys' MT cell activity was recorded while each monkey viewed arrays of moving dots, we would say the scientists intervened on the presentation of the array by beginning the computer program that the monkey watched. The array's appearing then caused activity in the monkey's eyeballs, which later caused activity in the MT cells. If the array had not appeared, other things constant, the monkey's MT cells would not have been active.

Bottom-up experiments are interventions on the subject's MT cells with respect to the subject's response of having perceived motion. In the Nichols and Newsome (2002) study, we can say the scientists intervened on the activity of the monkey's MT cells by stimulating them with an electrode while the monkey viewed the array, which later caused the monkey to mark its response by looking in a particular direction – namely, closer to the direction which the stimulated MT cells selectively respond to. In the Newsome and Paré (1988) study, we can say the scientists intervened on the MT cells by lesioning them, which then caused the monkey to give fewer correct answers about the direction of motion than before the lesions.

For any phenomenon, before executing the interventions, scientists likely will not know what lies on the causal chain between S_{IN} and S_{OUT}. The purpose of the interventions is to discover the components of S's ψ-ing. Scientists could begin with a vague functional and spatiotemporal characterization of the phenomenon. That is, they would need a sense of the spatiotemporal extension of S's ψ-ing, and of a few of its inputs and outputs. They would then execute the interventions to determine which parts' activities are on the causal chain that constitutes S's ψ-ing.

Alternatively, scientists might begin by discovering a collection of entities that causally interact, then identify a portion of the collection as a mechanism for some phenomenon. Which entities and activities count as components, instead of inputs or outputs, will depend on where they stipulate the spatiotemporal and functional boundaries of the mechanism to be. Their labelling some of the entities and activities components requires that they have an idea of the boundaries of the mechanism.

While this may sound like a problem for the causal situationist account, all accounts of constitutive relevance discussed here also demand that scientists know, more or less, the spatiotemporal boundaries of S's ψ-ing. Otherwise, they could not tell if they have caused changes in S's ψ-ing through interventions, nor if X is a physical part of S. Any account of constitutive relevance is likely to demand this, given what it is for something to be a component of a mechanism. The goal is to break an existing phenomenon down to its physical and active parts. Some knowledge of the existing phenomenon is necessary to proceed with this goal.

Unlike Baumgartner and Gebharter's (2015) account, which requires establishing that *every* cause of S's ψ-ing is also a cause of some X's Φ-ing, the causal situationist account does not demand empirical practices that are nearly impossible to achieve. In fact, the account is hardly more demanding than practices that establish causal relevance. Furthermore, we might say the causal situationist account is fulfilled whenever interlevel experiments successfully change both S's ψ-ing and X's Φ-ing with suitable control of confounding variables. These experiments are executions of ideal/ideal* interventions on S_{IN} with respect to X's Φ-ing and on X's Φ-ing with respect to S_{OUT}. The causal situationist account is fruitful empirically.

One might worry that satisfying the causal situationist account merely reveals a causal chain, rather than a mechanism. Weinberger (2017), for instance, argues that accounts of constitutive relevance that construe components as standing in causal relationships with S_{IN} and S_{OUT}, as the causal situationist account does, inadvertently reduce constitutive relevance to causal relevance. To secure constitutive relevance, which is an interlevel relationship, Weinberger says we would need a principled means of showing that S_{IN} and S_{OUT} lie on a different mechanistic level from Xs' Φ-ings. Otherwise, the ideal/ideal* interventions would establish mere intralevel causal relationships. This is problematic, he thinks, because these accounts say that the causal relationships between S_{IN}, X's Φ-ing, and S_{OUT} are the ones that make for the interlevel constitutive relationship between X's Φ-ing and S's ψ-ing. But if all we have shown are intralevel causal relationships, we do not get the interlevel constitutive relationship for free. Mechanistic explanations would, then, reduce to mere causal explanations.

The causal situationist account, first, does not reduce the constitutive relationship between some X's Φ-ing and S's ψ-ing to the causal relationships between S_{IN}, X's Φ-ing, and S_{OUT}. This is simply because it is not an account of what it is for something to be a component of a mechanism, metaphysically. Rather, it is an account of how to go about identifying components in practice. What it reduces is the empirical establishment of the constitutive relationship to the empirical establishment of certain causal and spatiotemporal relationships.

Second, the account has requirements that cannot be met by variables that are merely causally related. Suppose we wanted to show, by reductio, that A can count as a component of B by the causal situationist account, even though A is causally relevant to B. This is impossible. Given that A is causally relevant to B, it must not be the case that A is a physical part of B, since parts and wholes cannot causally interact. Therefore, A cannot count as a component of B by the causal situationist account, since it fails to fulfill criterion (CS1).

Suppose instead we wanted to show, by reductio, that, where A is causally relevant to B, B is causally relevant to C, and C is causally relevant to D, the causal situationist account can mistakenly treat this causal chain as a mechanism and identify one of the variables as a component. This too is impossible. Criteria (CS2) and (CS3) could be met; an ideal intervention on A can change B, and an ideal* intervention on B can change C. But B still cannot count as a component of D by the causal situationist account. Given that B and D are causally related, B must not be a physical part of D, violating (CS1). Or, D does not cause C, whereas in an actual mechanism, S's ψ-ing causes S_{OUT}, by definition, so applying the causal situationist account is inappropriate to begin with. In either case, variables that stand in mere causal relationships cannot fulfill the causal situationist account. Variables that stand in constitutive relationships can fulfill the account. This suggests that the causal situationist account does not, and cannot, establish mere causal relationships.

Third, we can maintain that every X's Φ-ing is on a different level from S's ψ-ing. All it is for some variable, A, to be on a lower mechanistic level than some variable, B, is for A to be a component of B (Povich and Craver 2017; Craver and Bechtel 2007). Because the causal situationist account allows for the identification of components, it thereby enables us to say that some things are on lower or higher mechanistic levels than other things. X's Φ-ing being on a lower level than S's ψ-ing does not require that it is also on a lower level from S_{IN} and S_{OUT}. X's Φ-ing is not the kind of thing that could be on a lower (nor higher) level than S_{IN} and S_{OUT}, since it is not related to them constitutively (see, e.g., Povich and Craver 2017). There is no question, then, of an interlevel relationship between X's Φ-ing and S_{IN} and S_{OUT}.

Even if it is the case, as I have granted, that S_{IN} and S_{OUT} are defined by their causal relationships to S's ψ-ing, and at least some inputs and outputs must be identified before any components of S's ψ-ing can be identified, this does not mean that S_{IN} and S_{OUT} indeed lie on the same mechanistic level as S's ψ-ing. Any X's Φ-ing is on a lower level than S's ψ-ing, and the causal situationist account establishes causal relationships between S_{IN}, X's Φ-ing, and S_{OUT}. To refrain from admitting interlevel causation, then, we cannot say that S_{IN} and S_{OUT} are on the same level as S's ψ-ing. We can maintain, nevertheless, that Xs' Φ-ings are on a lower level than S's ψ-ing just because they constitute S's ψ-ing.

On the causal situationist account, then, the general term 'interlevel experiment' and the specific terms 'top-down experiment' and 'bottom-up experiment' are not quite appropriate. The relevant interventions do not involve variables that stand on different mechanistic levels.

Even though satisfaction of the causal situationist account does demonstrate causal relationships between S_{IN}, X's Φ-ing, and S_{OUT}, these are not the only relationships the account demands. It also requires a spatiotemporal relationship between X and S, and causal relationships between S_{IN}, S's ψ-ing, and S_{OUT}. This bundle of relationships, rather than just the causal relationships Weinberger discusses, is what makes for constitutive relevance between X's Φ-ing and S's ψ-ing. X's being a physical part of S, and being caused by and causing the same things as S's ψ-ing, are what entail that X's Φ-ing is a component of S's ψ-ing.

It is also due to this bundle of relationships that scientists cannot lose sight of S's ψ-ing when using the causal situationist account. It would not be possible, as long as the criteria are in fact being met, to accidentally highlight nothing but causal relationships. Nor, for that matter, to highlight relationships that are utterly irrelevant to S's ψ-ing. Working with variables that scientists already know are causally and spatiotemporally related to S's ψ-ing and doing the relevant interventions ensure that they are discovering the actual components of S's ψ-ing.

To summarize the causal situationist account, it enables scientists to discover the components of a mechanism by situating components on the causal chain between S_{IN} and S_{OUT}. Because S's ψ-ing is that which causally occurs between S_{IN} and S_{OUT}, demonstrating that X is a part of S and is causally related to S_{IN} and S_{OUT} shows that X's Φ-ing helps constitute S's ψ-ing. It does not indicate interlevel causation. It therefore satisfies the three constraints very simply; it requires little more than thinking about the interventions executed in interlevel experiments as occurring on S_{IN} with respect to X's Φ-ing and on X's Φ-ing with respect to S_{OUT}.

5 Conclusion

Contra Baumgartner and Casini (2017), it is possible, then, to create an account of constitutive relevance that respects the three constraints from Craver (2007). The problem, it seems, is not with trying to build an account that satisfies those three constraints, but with which variables we take the interventions executed in interlevel experiments to occur between. The causal situationist account straightforwardly fits into Craver's framework. If indeed the three constraints are accurate, the causal situationist account enables us to make sense of how some of the empirical practices that scientists actually use establish constitutive relevance, without misconstruing the constitutive relationship as causal. It is thus fruitful empirically and attractive theoretically and descriptively.

Acknowledgements Thanks to Carl Craver and Ben Henke for their great help developing this paper, as well as to the audience at the 2017 Meeting of the Society for the Metaphysics of Science in New York.

References

Baumgartner, M., & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science*, 84, 214-233.

Baumgartner, M., & Gebharter, A. (2015). Constitutive relevance, mutual manipulability, and fat-handedness. *British Journal for the Philosophy of Science*, 67, 731-756.

Bechtel, W. (In press). The epistemology of evidence in cognitive neuroscience. In R. Skipper, C. Allen, R. A. Ankeny, C. F. Craver, L. Darden, G. Mikkelson, & R. Richardson (Eds.), *Philosophy across the life sciences*. Cambridge: MIT Press.

Bechtel, W., & Abrahamsen, A. (2005) Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421-441.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, 12(12), 4745-4765.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22(4), 547-563.

Gebharter, A. (2017). Uncovering constitutive relevance relations in mechanisms. *Philosophical Studies*, 174(11), 2645-2666.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(S3), S342-353.

Harbecke, J. (2015). The regularity theory of mechanistic constitution and a methodology for constitutive inference. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 54, 10-19.

Harinen, T. (2018). Mutual manipulability and causal inbetweenness. *Synthese*, 195(1), 35-54.

Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology & Philosophy*, 27(4), 545-570.

Krickel, B. (2018). Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science Part A*, 68, 58-67.

Newsome, W. T., & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, 8(6), 2201-2211.

Nichols, M. J., & Newsome, W. T. (2002). Middle temporal visual area microstimulation influences veridical judgments of motion direction. *Journal of Neuroscience*, 22(21), 9530-9540.

Povich, M., & Craver, C. F. (2017). Mechanistic levels, reduction, and emergence. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 185-197). New York: Routledge.

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese*, 192(11), 3731-3755.

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346(6280), 174-177.

Weinberger, N. (2017). Mechanisms without mechanistic explanation. *Synthese*, https://doi.org/10.1007/s11229-017-1538-1

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings*. Harvard: Harvard University Press.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91(2), 303-347.