CONSTITUTION AND CAUSAL ROLES

Lorenzo Casini[†] and Michael Baumgartner[‡]

Alexander Gebharter (2017b) has proposed to use Bayesian network (BN) causal discovery methods to identify the constitutive dependencies underwriting mechanistic explanations. The account assumes that mechanistic constitution behaves like deterministic direct causation, such that BN methods are directly applicable to mixed variable sets featuring both causal and constitutive dependencies. Gebharter claims that such mixed sets, under certain restrictions, comply with the assumptions of the causal BN framework. The aim of this paper is twofold. In the first half, we argue that Gebharter's proposal incurs severe problems, ultimately rooted in widespread non-compliance of mechanistic systems with BN assumptions. In the second half, we present an alternative way to bring BN tools to bear on the discovery of mechanistic constitution. More precisely, we argue that all of a phenomenon's parts, whose causal roles account for why the phenomenon has its characteristic causal role, are constituents—where the notion of causal role is probabilistically understood.

1 INTRODUCTION

The mechanistic account of scientific explanation (Machamer et al., 2000; Bechtel and Abrahamsen, 2005; Glennan, 2002) holds that the explanandum, a higher-level phenomenon, is explained by the lower-level mechanism responsible for it. In a popular characterization,

> [a] mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena. (Bechtel and Abrahamsen, 2005, 423)

To give a simple but paradigmatic example, which shall serve as our guiding example throughout the paper, the phenomenon of amplification in a two-stage amplifier is caused by a signal (e.g., current, voltage, power) received from an input source, and causes effects such as signal distortion in an output device (e.g., a loudspeaker). The phenomenon is explained by the augmentation of the signal by the amplifier's two transistors arranged in series (see Wimsatt 2007, ch. 12).

[†]Dept. of Philosophy, University of Geneva. Email: lorenzo.casini@unige.ch
[‡]Dept. of Philosophy, University of Bergen. Email: michael.baumgartner@uib.no

More generally, a mechanism is embedded in a causal context, where causal background conditions are operative relative to which certain parts of the system are responsible for the phenomenon. The relevant kind of responsibility is *constitutive* rather than causal. The system's parts that mechanistically explain the phenomenon are the "component" (cf. quote), or *constituent*, parts. While causation has been at the centre of philosophical theorizing for centuries, the notion of constitution, or constitutive relevance, has only recently begun to attract philosophical attention. In particular, it is still unclear what discovery method(s) could systematize the data-based inference to constitution.

Gebharter (2017b) has suggested drawing on the resources of the Bayesian network (BN) framework, which is widely used to model and discover causation (Spirtes et al., 2000; Pearl, 2000), to address the task of constitutive discovery. He claims that, despite the differences between causation and constitution, the BN axioms used to model causation also capture constitution, and that constitution can be implicitly characterized as a form of deterministic direct causation. He concludes that BN causal discovery algorithms—PC, in particular—may concurrently be used for both causal and constitutive discovery.

After a brief introduction to causal BNs (§2), the first part of this paper (§3) takes issue with this latter conclusion. Variable sets processed by BN procedures must satisfy specific assumptions. Violations of these assumptions have been argued to be rare in variable sets exclusively featuring causal relations, which are assumed to be non-deterministic (or pseudoindeterministic) in the BN framework. Therefore, BN assumptions may be justifiably assumed for causal contexts. Constitutive relations, by contrast, generate deterministic dependencies, in the presence of which violations of BN assumptions are no longer rare but commonplace, which undermines their justifiable assumability. Moreover, we argue that no systematically reliable inferences can be drawn outside the scope of validity of those assumptions. This latter point is illustrated and substantiated by a series of inverse search trials involving data simulations, which evaluate the performance of the PC algorithm when applied to mechanistic systems.

The second part of the paper (§4) proposes a sufficient condition for constitution that avoids these problems and allows for bringing BN methods to bear on the task of constitutive discovery in a theoretically sound way. In a nutshell, our proposal is that all of a phenomenon's parts, whose causal roles account for why the phenomenon has its characteristic causal role, are constituents. This idea has been recently expressed, in one way or another, by a number of authors (e.g., Gillett 2002, 319[1]; see also Fazekas and Kertész 2011 and Soom 2012[2])

[1] Gillett (2002) proposes an account of "realization" as a relation between the causal powers individuating a phenomenon and those individuating its constituents.

[2] Contrary to Gillett (2002), Fazekas and Kertész (2011) and Soom (2012) maintain that the causal role of the phenomenon is identical to the causal role of its constituents. We do not endorse this assumption.

but it has never been cashed out in detail and with formal precision. We fill this gap by giving it a precise rendering in the BN framework, which is particularly suitable to explicate the notion of causal roles figuring in our account. We also demonstrate the performance of the proposal by means of a series of inverse search trials.

2 PRELIMINARIES

We begin by introducing the theory of causal BNs, as well as a notational convention on the variables of BNs representing mechanistic systems.

Traditionally, the BN formalism uses generic random variables to represent types (or degrees) of properties or behaviours independently of the entities instantiating them. Here, however, we shall follow the mechanistic literature in taking the variables as denoting the behaviours exhibited by specific entities (such as a system and its constituents), and consequently adopt the following notational convention. Calligraphic fonts are used for *specific* random variables $\mathcal{A}(\mathrm{S})$ and $\mathcal{B}(\mathrm{P}_1)$ (Spohn, 2006), by which we denote the behaviour $\mathcal{A}$ of a specific system S and the behaviour $\mathcal{B}$ of a specific part P_1. As we are only concerned with specific variables, we will leave the entity-relativity of our variables implicit and just write "$\mathcal{A}$", "$\mathcal{B}$", etc. for the behaviour types "$\mathcal{A}(\mathrm{S})$", "$\mathcal{B}(\mathrm{P}_1)$", etc.

A BN is a triple $\langle \mathbf{V}, \mathbf{E}, \Pr \rangle$ of a finite set $\mathbf{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_n\}$ of variables, each taking finitely many possible values; a set of edges $\mathbf{E}$ over the variables in $\mathbf{V}$, such that variables and edges $\langle \mathbf{V}, \mathbf{E} \rangle$ form a directed acyclic graph (DAG); and a probability distribution $\Pr$, such that the probability of each variable $\mathcal{V}_i$ in the DAG obeys the Markov Condition (MC):

(MC) For any $\mathcal{V}_i \in \mathbf{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_n\}$, $\mathcal{V}_i \perp\!\!\!\perp \mathbf{Non}_i \mid \mathbf{Par}_i$,

where $\mathbf{Par}_i$ denotes the set of parents of $\mathcal{V}_i$, and $\mathbf{Non}_i$ denotes the set of non-descendants of $\mathcal{V}_i$.[3] In words, each variable is probabilistically independent of its non-descendants, conditional on its parents. For instance, MC applied to the DAG in Figure 1 implies that $\mathcal{V}_4 \perp\!\!\!\perp \mathcal{V}_1, \mathcal{V}_5 \mid \mathcal{V}_2, \mathcal{V}_3$. In BN jargon, $\mathcal{V}_2$ and $\mathcal{V}_3$ *screen off* $\mathcal{V}_4$ from $\mathcal{V}_1$ and $\mathcal{V}_5$.

If a BN is causally interpreted, the edges stand for direct causal relations, $\mathbf{Par}_i$ denotes the set of direct causes of $\mathcal{V}_i$, $\mathbf{Non}_i$ the set of $\mathcal{V}_i$'s non-effects in the true causal structure regulating the behaviour of the variables in $\mathbf{V}$, and MC is called Causal Markov Condition (CMC) (cf. Spirtes et al. 2000, §3.4.1, §3.5.1).

In addition, the Causal Faithfulness Condition (CFC) is often assumed in the causal BN literature (Zhang and Spirtes 2008, 247):

[3] The "parents" of $\mathcal{V}_i$ are the direct ancestors of $\mathcal{V}_i$, namely those variables on directed paths into $\mathcal{V}_i$ from which $\mathcal{V}_i$ can be reached without mediation via other variables. The "descendants" of $\mathcal{V}_i$ are those variables, including $\mathcal{V}_i$, which may be reached from $\mathcal{V}_i$ along a directed path.

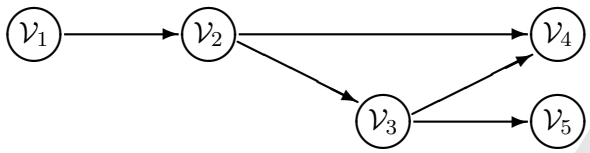

Figure 1: A Bayesian network

(CFC) $\langle \mathbf{V}, \mathbf{E}, \Pr \rangle$ is such that every conditional independence relation true in $\Pr$ is entailed by CMC applied to the true DAG $\langle \mathbf{V}, \mathbf{E} \rangle$.

CFC guarantees that there is no causal dependence without probabilistic dependence—i.e., the only probabilistic independencies in the graph are due to the absence of causal dependencies. For instance, CFC applied to the BN in Figure 1 implies $\mathcal{V}_4 \not\perp\!\!\!\perp \mathcal{V}_2$, that is, there is no exact cancellation of $\mathcal{V}_2$'s effect on $\mathcal{V}_4$ due to, say, a positive influence along the direct path $\mathcal{V}_2 \longrightarrow \mathcal{V}_4$ and a negative influence along the indirect path $\mathcal{V}_2 \longrightarrow \mathcal{V}_3 \longrightarrow \mathcal{V}_4$. In particular, CFC entails that all causal dependencies are detectable by conditional independence tests, as commonly performed by BN constraint-based algorithms for causal discovery (see, e.g., Spirtes et al. 2000, 82, 88, 144, and Pearl 2000, 50, 52).

A wide array of algorithms (e.g., SGS, PC, FCI, IC) are used under the assumptions of CMC and CFC, as CMC and CFC are sufficient for the correctness of these algorithms. In many well-known contexts, CMC and CFC are provably satisfied or only rarely violated, such that these assumptions are justified, to the effect that algorithms assuming CMC and CFC are reliable in those contexts.

On the one hand, CMC is provably satisfied if (i) the functional relations in the data-generating structure are linear, (ii) the exogenous variables and error terms are independently distributed, (iii) all non-deterministic dependencies in the data (i.e. dependencies not producing conditional probabilities equal to 1) are due to noise and not to some fundamentally indeterministic process, that is, all non-deterministic dependencies are so-called *pseudoindeterministic*, and (iv) the variable set is *causally sufficient*, where (causal) Sufficiency is defined as follows (Zhang 2006, 8; cf. Spirtes et al. 2000, §3.2.2):

(Sufficiency) $\langle \mathbf{V}, \mathbf{E}, \Pr \rangle$ is such that every direct common cause of any two variables in $\mathbf{V}$ either is in $\mathbf{V}$ or has an ancestor in $\mathbf{V}$ or has the same value for all units in the population.

Sufficiency guarantees that for any two variables in $\mathbf{V}$, there is no probabilistic dependence not due to a causal dependence—i.e., no probabilistic dependence is spurious.

On the other hand, CFC holds if (i) and (ii) hold, and (v) the data do not contain deterministic but only pseudoindeterministic dependencies. In that case,

violations of CFC have Lebesgue measure 0, which entails that they are very rare (Spirtes et al., 2000, §3.5).

At the same time, there are well-known contexts in which BN assumptions are frequently violated and, hence, not justified. One such context that will be particularly relevant for the remainder of this paper consists in the presence of deterministic dependencies (which generate conditional probabilities equal to 1) in the data. Given determinism, violations of CFC are commonplace (Spirtes et al. 2000, §3.8; Glymour 2007, 236). To illustrate, assume that the dependencies along the path $\mathcal{V}_1 \longrightarrow \mathcal{V}_2 \longrightarrow \mathcal{V}_4$ in the causal structure of Figure 1 are deterministic, meaning that $\mathcal{V}_1$ determines $\mathcal{V}_2$, which determines $\mathcal{V}_4$. It then holds that $\Pr(\mathcal{V}_4 \,|\, \mathcal{V}_1 \wedge \mathcal{V}_2) = \Pr(\mathcal{V}_4 \,|\, \mathcal{V}_1) = 1$, *viz.* that the indirect cause $\mathcal{V}_1$ screens off $\mathcal{V}_4$ from its direct cause $\mathcal{V}_2$, which, however, is not entailed by CMC and hence violates CFC. That CFC violation does not hinge on the particularities of the BN in Figure 1 but generalizes: *all* (and not just some peculiar and rare) deterministic chains violate CFC. In light of the frequency of CFC violations under determinism, standard BN algorithms are normally assumed to be inapplicable to deterministic data.

3 THE LIMITS OF GEBHARTER'S PROPOSAL

3.1 The proposal

While BNs have a long tradition of successful applications in causal discovery, they have played no role so far in constitutive discovery. The main reason is that constitution is commonly assumed to be characterized by (non-reductive) *supervenience* (see, e.g., Glennan 1996, 61-2, and Eronen 2011, ch. 11), which generates deterministic dependencies: a complete set of constituents of a phenomenon amounts to a supervenience base of that phenomenon, and thus a sufficient condition for it—i.e., necessarily, there is no change in the phenomenon without a change in its supervenience base. By contrast, as indicated in the previous section, standard BN algorithms are typically only applied to indeterministic data, which, moreover, are assumed to be pseudoindeterministic.

To further clarify the difference between pseudoindeterministic and deterministic dependencies, consider the mechanism operating in an amplifier. Let $\mathcal{G}$ represent the phenomenon of gain, or total voltage increase, of an amplifier subject to a voltage input $\mathcal{I}$. Amplifiers are built by assembling a number of active elements, usually transistors, in a circuit. We assume that the amplifier in question is a two-stage amplifier, such that the signal received by a first transistor is amplified and fed to a second transistor, which further amplifies it. Let $\mathcal{A}$ and $\mathcal{B}$ be the transistors' individual gains. Then, the amplifier's overall gain in response to any given input $\mathcal{I} = i$ is some pseudoindeterministic function $\mathcal{G} = r_{\mathcal{G}} i - i + \epsilon_{\mathcal{G}}$, where $r_{\mathcal{G}}$ indicates the amplifiers amplification ratio and $\epsilon_{\mathcal{G}}$ is a noise term. For instance, if $\mathcal{I} = 2$ volts and the amplification ratio is 8, then the overall gain is $\mathcal{G} = 2 \times 8 - 2 + \epsilon_{\mathcal{G}}$ volts, where 14 (i.e., $2 \times 8 - 2$) volts and

$\epsilon_{\mathcal{G}}$ volts, respectively, are $\mathcal{G}$'s deterministic and non-deterministic components. Analogously, the transistors' gains are given by $\mathcal{A} = r_{\mathcal{A}}i - i + \epsilon_{\mathcal{A}}$ volts and $\mathcal{B} = r_{\mathcal{B}}i - i + \epsilon_{\mathcal{B}}$ volts. The amplification ratio of a serial amplifier is—ideally, i.e., ignoring the role of the amplifier's passive components—the product of its transistors' amplification ratios. Assume that the first transistor amplifies by a ratio 2, and the second amplifies by a ratio 4, such that the amplifier's ratio is 8. Then, when subject to an input $\mathcal{I} = 2$ volts, the first transistor amplifies the signal by a gain of $2 \times 2 - 2 + \epsilon_{\mathcal{A}}$ volts; and the second transistor receives that signal and amplifies it further by a gain of $4 \times (4 + \epsilon_{\mathcal{A}}) - (4 + \epsilon_{\mathcal{A}}) + \epsilon_{\mathcal{B}}$ volts. By contrast, the relation between overall gain $\mathcal{G}$ on the one hand, and the transistors' individual gains $\mathcal{A}$ and $\mathcal{B}$ on the other hand, is not pseudoindeterministic but deterministic: $\mathcal{G}$ is simply the sum of $\mathcal{A}$ and $\mathcal{B}$, meaning that $\mathcal{A}$ and $\mathcal{B}$ determine $\mathcal{G}$, such that whatever noisy component is present in $\mathcal{G}$, it is inherited from, and fully accounted for by, the noise in $\mathcal{A}$ and $\mathcal{B}$. More precisely, supervenience entails that $r_{\mathcal{G}}i - i + \epsilon_{\mathcal{G}} = r_{\mathcal{B}}(r_{\mathcal{A}}i + \epsilon_{\mathcal{A}}) - i + \epsilon_{\mathcal{B}}$. When $\mathcal{I} = 2$ volts, $2 \times 8 - 2 + \epsilon_{\mathcal{G}} = 4 \times (4 + \epsilon_{\mathcal{A}}) - 2 + \epsilon_{\mathcal{B}}$, that is, $\epsilon_{\mathcal{G}} = 4\epsilon_{\mathcal{A}} + \epsilon_{\mathcal{B}}$.

Notwithstanding the frequency of CFC violations under determinism, Gebharter (2017b, 2652–54) has—surprisingly—argued that constitution satisfies the same properties that the BN framework assumes for causation. More specifically, he contends that the screening-off behaviour of complete sets of constituents (i.e. sets comprising a complete supervenience base of a phenomenon) is analogous to that of deterministic direct causes and that the screening-off behaviour of incomplete sets is analogous to that of indeterministic direct causes. From that, he infers that constitutive relations can be represented by BNs and that, with some restrictions, BN algorithms can be directly applied to variable sets featuring both constitutive and causal relations, such that the uncovered dependencies can then be grouped into causal and constitutive dependencies using information about spatiotemporal overlap (i.e. about parthood relations). In short, Gebharter claims that BN algorithms—in particular, PC —can be used to perform causal and constitutive search in one go.

Given the well-known problems determinism creates for BN algorithms, the natural conclusion to draw from Gebharter's finding that constitution behaves like deterministic direct causation would be that BNs are *not* capable of representing systems featuring constitutive relations and that—*a fortiori*—BN algorithms are not applicable to systems featuring constitutive relations. Gebharter is aware that his proposal raises severe questions. He discusses two conceivable approaches to reconcile the deterministic nature of constitution with BN assumptions (cf. Gebharter 2017b, 2661–62):

(A) Only apply BN algorithms to incomplete constitutive sets, which do not amount to complete supervenience bases and, hence, do not generate deterministic dependencies in the first place;

(B) Allow for deterministic dependencies but only apply BN algorithms to systems featuring no more than two mechanistic levels.

Approach (A) amounts to testing for deterministic dependencies prior to a BN analysis (by, e.g., performing a multicollinearity test) and, if that test is positive, abstain from applying BN algorithms. A variable set $\mathbf{V}$ featuring constitutive relations will only be free of deterministic dependencies provided that no phenomenon in $\mathbf{V}$ has a complete set of constituents in $\mathbf{V}$. As constitution, according to Gebharter, technically behave like causation, missing constituents are on a par with missing causes of the phenomenon. Since constituents typically are not only relevant for the phenomenon but also for other micro-level variables contained in $\mathbf{V}$, it follows that missing constituents amount to missing common causes of variables in $\mathbf{V}$, in violation of causal Sufficiency (Gebharter, 2017b, 2660). Yet, if Sufficiency is violated, CMC tends to be violated as well. Since adopting approach (A) in an attempt to avoid CFC violations generates frequent violations of CMC, Gebharter concludes that it fails to reconcile the deterministic nature of constitution with BN assumptions. To justifiably assume CMC, $\mathbf{V}$ should contain complete constitutive sets, meaning that data over $\mathbf{V}$ should feature deterministic dependencies.

This leaves us with approach (B), which Gebharter indeed advances as a solution to the problems prompted by the deterministic nature of constitution (Gebharter, 2017b, 2662). In the previous section, we have seen that chains of at least three deterministically related variables are a paradigmatic type of structure generating CFC violations. Without argument, Gebharter takes such chains to be the source of *all* CFC violations induced by determinism. Accordingly, he stipulates that BN algorithms only be applied to mechanistic systems with no more than two levels, which excludes deterministic chains. More specifically, Gebharter proposes to use background knowledge on spatiotemporal parthood relations between instances of analysed variables in order to only include parts of the phenomenon in $\mathbf{V}$ but not parts *of parts of* the phenomenon. That is, $\mathbf{V}$ must not contain any triple of variables $\langle \mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3 \rangle$ such that $\mathcal{V}_1$ is a part of $\mathcal{V}_2$, which is a part of $\mathcal{V}_3$.[4] Gebharter believes that this two-level restriction ensures that deterministic dependencies do not conflict with CFC more frequently than pseudoindeterministic dependencies and, hence, that CFC is justifiably assumable even for the purpose of constitutive discovery.

3.2 Extensive Faithfulness violations

Gebarther severely underestimates the problems constitutive relations induce for BN algorithms. First, recall that, in order to justifiably assume CMC (and Sufficiency), every analysed variable set $\mathbf{V}$ should contain a complete set of constituents $\mathbf{C}$ for every phenomenon in $\mathbf{V}$. Subject to the (presupposed) supervenience of phenomena on their constituents, every phenomenon is deter-

[4] In the interest of brevity, we speak deliberately loosely—here and in the remainder of the paper—about variables being related by spatiotemporal parthood. Of course, it is not the variables themselves that are related by spatiotemporal parthood but the entities represented by these variables.

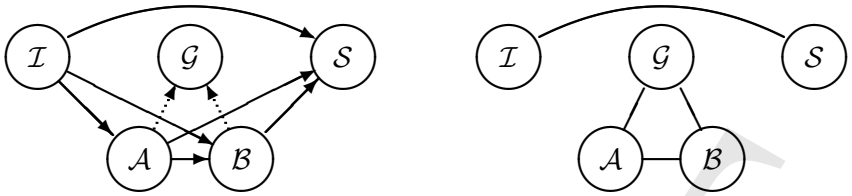

Figure 2: (a) Structure of a two-stage amplifier mechanism over $\mathbf{G} = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}\}$, for an epiphenomenalist*. Dotted arrows are constitutive; all other arrows are causal. (b) Graph over $\mathbf{G}$ where all edges that can be screened off have been eliminated.

mined by $\mathbf{C}$. This *bottom-up* determination yields that every phenomenon is screened off from all other variables—whether in $\mathbf{V}$ or not. The reason is that determination is monotonic: if $\mathbf{C}$ determines $\mathcal{V}_1$, then $\mathbf{C} \wedge \mathcal{V}_i$ also determines $\mathcal{V}_1$, where $\mathcal{V}_i$ is an arbitrary variable. If $\Pr(\mathcal{V}_1|\mathbf{C}) = 1$, then $\Pr(\mathcal{V}_1|\mathbf{C} \wedge \mathcal{V}_i) = 1$, meaning that $\mathbf{C}$ screens off $\mathcal{V}_1$ from any variable $\mathcal{V}_i$. Accordingly, in CMC-warranting contexts, every mechanistic system (involving two or more levels) will feature probabilistic independencies between phenomena and all their non-constituents. Likewise assuming CFC in such contexts would imply that these independencies are entailed by the true graphs, meaning that all macro phenomena are both uncaused and causally inert, that is, *causally isolated*.

However, as most mechanists—the addressees of Gebharter's proposal—endorse the existence of macro-level causation,[5] they will reject the inference to the causal isolation of all phenomena by not assuming CFC in CMC-warranting contexts. Instead, they will interpret the independencies between phenomena and all non-constituents as yet another CFC violation induced by determinism. Clearly, this breach of CFC obtains even in two-level systems. To illustrate, reconsider our amplifier example and let the analysed variable set be $\mathbf{G} = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}\}$, where $\mathcal{I}$ (the amplifier's input), $\mathcal{G}$ (the amplifier's overall gain), $\mathcal{A}$ (the first transistor's gain), and $\mathcal{B}$ (the second transistor's gain) are complemented by $\mathcal{S}$, which denotes, say, the distortion of the signal as received by a loudspeaker. Since $\mathcal{A}$ and $\mathcal{B}$ determine $\mathcal{G}$ (from the bottom up), $\mathcal{A}$ and $\mathcal{B}$ screen $\mathcal{G}$ off from $\mathcal{I}$ and $\mathcal{S}$, or formally $\mathcal{I}, \mathcal{S} \perp\!\!\!\perp \mathcal{G} \,|\, \mathcal{A}, \mathcal{B}$. But if input and distortion are a cause and an effect of an amplifier's gain in the true graph, these conditional independencies violate CFC.

Such bottom-up determination can be reconciled with CFC by rejecting that input and distortion actually are a cause and an effect of an amplifier's gain; more generally, by rejecting the possibility of macro-level causation, that is, by committing to the metaphysical view that phenomena really are causally isolated. That amounts to a radical form of macro-level epiphenomenalism, call it *epiphenomenalism**, *viz.* the view that non-fundamental properties are

[5] In fact, we are not aware of a single proponent of the mechanistic framework who would endorse the causal isolation of macro phenomena.

not only causally inert (as entailed by standard epiphenomenalism) but also uncaused. More concretely, according to epiphenomenalism*, the true graph for our amplifier example is the one in Figure 2a. Against that background, the fact that $\mathcal{G}$ is screened off from $\mathcal{I}$ and $\mathcal{S}$ by $\mathcal{A}$ and $\mathcal{B}$ follows from CMC applied to the true graph and, hence, does not violate CFC.

To uphold his claim that two-level systems do not violate CFC, Gebharter indeed endorses epiphenomenalism* (cf. Gebharter, 2017a). While this manoeuvre reconciles bottom-up determination with CFC, it clashes with the metaphysical commitments of most mechanists. Despite a longstanding debate among philosophers on whether the notion of causation is dispensable in fundamental physics (Russell, 1913; Norton, 2003; Frisch, 2012), it is uncontroversial that macro-level disciplines, such as the social and biomedical sciences, routinely engage in investigating causal relations among macro variables and, hence, do not commit to epiphenomenalism*. A characterization of constitution that is at odds with the scientific practice of many non-fundamental disciplines is, at best, undesirable. This holds all the more in view of the fact that dependencies between macro variables pass the usual BN tests for causation in variable sets without variables in parthood and supervenience relations—tests the validity of which Gebharter does not dispute.

What is worse, allowing deterministic dependencies in data processed by common BN algorithms generates further problems, which—contrary to epiphenomenalism*—Gebharter cannot possibly accept. In particular, in addition to bottom-up determination mechanistic systems with exactly two levels may also feature *top-down* determination, to the effect that not only phenomena are screened off from all incoming and outgoing influences, but also constituents can be screened off in this way. This problem is best introduced by reconsidering the amplifier example. The amplifier's absolute overall gain $\mathcal{G}$ is the sum of its constituents $\mathcal{A}$ and $\mathcal{B}$. The function of addition, however, is reversible: it not only holds that $\mathcal{G}$ is determined by $\mathcal{A}$ and $\mathcal{B}$, but also that $\mathcal{A}$ is determined by $\mathcal{G}$ and $\mathcal{B}$ (e.g., $\mathcal{G}=14 \wedge \mathcal{B}=12$ determines $\mathcal{A}=2$) and that $\mathcal{B}$ is determined by $\mathcal{G}$ and $\mathcal{A}$ (e.g., $\mathcal{G}=14 \wedge \mathcal{A}=2$ determines $\mathcal{B}=12$). Hence, every variable in the set $\mathbf{M} = \{\mathcal{G}, \mathcal{A}, \mathcal{B}\}$ is screened off from $\mathcal{I}$ and $\mathcal{S}$ by the other two elements of $\mathbf{M}$. To illustrate, all adjacencies corresponding to these conditional independencies are removed from the graph skeleton in Figure 2b. A causal/constitutive interpretation of that graph entails that the mechanism over $\mathbf{M}$ is causally isolated from its environment.

This shows that, even if we grant Gebharter his epiphenomenalism*, our amplifier still violates CFC. Subject to epiphenomenalism*, the true graph over $\mathbf{G}$ is the one in Figure 2a. Hence, the top-down determination induced by the reversible functional dependencies among the elements of $\mathbf{M}$ generates additional conditional independencies not entailed by CMC applied to the true graph. These additional independencies violate CFC. That is, in mechanistic systems featuring no more than two levels, CFC may be violated even against

the backdrop of epiphenomenalism*. To avoid this consequence, Gebharter would have to endorse the absence of causal influences not only in and out of $\mathcal{G}$ but also in and out of $\mathcal{A}$ and $\mathcal{B}$. That is, he would have to contend that the mechanisms responsible for the gains of amplifiers are causally isolated from the rest of the universe. We take it as a given that Gebharter is not prepared to go that far.

The crucial follow-up question now becomes how widespread top-down determination is in mechanistic systems. It is clearly not limited to amplifier gains or even to phenomena whose values are the sum of their constituents. Top-down determination obtains whenever the relation between phenomena and their constituents is regulated by an aggregation function with the following *reversibility property*: a function $y = f(x_1, \ldots, x_n)$ is reversible iff all of its inputs x_i are determined by its output y in conjunction with all of its other inputs apart from x_i, or formally, iff for all i, $1 \leq i \leq n$, $x_i = f^{-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n, y)$. Examples of functions for which reversibility holds are linear functions, the product of non-zero values, exponentiation of positive integers, the sum of squares, many Boolean functions, or functions used in information coding, storage, and encryption (which are explicitly exploited for their reversibility).

To provide another example, consider the phenomenon of voting by a show of hands. Casting a vote, $\mathcal{W} = 1$, can be constituted by a raise of either the left hand, $\mathcal{L} = 1$, or of the right hand, $\mathcal{R} = 1$ (but raising both hands is invalid); or formally, $\mathcal{W} = 1 \leftrightarrow (\mathcal{L} = 1 \wedge \mathcal{R} = 0) \vee (\mathcal{L} = 0 \wedge \mathcal{R} = 1)$. This system of binary variables does not only feature bottom-up determination but also top-down determination: any of the four possible value configurations of $\{\mathcal{W}, \mathcal{L}\}$ and of $\{\mathcal{W}, \mathcal{R}\}$ determine the value of $\mathcal{R}$ and $\mathcal{L}$, respectively.[6] Hence, not only the phenomenon of voting but also the hand raisings are screened off from all variables outside of that system. But clearly, outside variables can de facto causally interact with hand raisings (e.g. they have causes in the motor cortex and effects in air displacement), which, in turn, entails that these conditional independencies violate CFC.

These considerations suffice to establish that, contrary to what Gebharter envisages in approach (B), CFC violations in (deterministic) mechanistic systems comprising only two levels are not rare but widespread—unlike CFC violations in (pseudoindeterministic) causal systems. A possible response to this objection might be to further restrict the applicability of BN methods in constitutive inference. More concretely, Gebharter could impose that his procedure is only applicable to two-level systems satisfying the additional constraint that the behaviours of the phenomenon and its constituents are regulated by a *non*-reversible aggregation function. Paradigmatic non-reversible functions are, for instance, periodic functions, products of zero, or the maximum and minimum functions. If a phenomenon is aggregated from its constituents by a

[6] To illustrate for $\{\mathcal{W}, \mathcal{L}\}$ and $\mathcal{R}$: $\mathcal{W} = 0 \wedge \mathcal{R} = 0 \rightarrow \mathcal{L} = 1$; $\mathcal{W} = 0 \wedge \mathcal{R} = 1 \rightarrow \mathcal{L} = 0$; $\mathcal{W} = 1 \wedge \mathcal{R} = 0 \rightarrow \mathcal{L} = 1$; and $\mathcal{W} = 1 \wedge \mathcal{R} = 1 \rightarrow \mathcal{L} = 0$.

non-reversible function, it does not hold for every constituent that its values are determined by the phenomenon in conjunction with all other constituents, that is, top-down determination does not obtain.

However, restricting the applicability of Gebharter's procedure to systems regulated by certain types of aggregation functions differs in a crucial respect from Gebharter's original approach (B), which only restricts it to two-level systems. A necessary condition for two variables $\mathcal{V}_1$ and $\mathcal{V}_2$ to be located at different mechanistic levels is that either $\mathcal{V}_1$ is a spatiotemporal part of $\mathcal{V}_2$ or that $\mathcal{V}_2$ is a spatiotemporal part of $\mathcal{V}_1$. Hence, a variable set $\mathbf{V}$ can be said to contain variables of no more than two mechanistic levels if it does not contain a triple $\langle \mathcal{V}_i, \mathcal{V}_j, \mathcal{V}_k \rangle$ such that $\mathcal{V}_i$ is a part of $\mathcal{V}_j$ and $\mathcal{V}_j$ is a part of $\mathcal{V}_k$. While identifying spatiotemporal parthood relations—clarity on which is generally assumed in the mechanistic literature—is undoubtedly difficult, it does not presuppose clarity on constitutive relations. In consequence, that $\mathbf{V}$ satisfies the two-level restriction can be established *independently* of clarity on the constitutive relations among the elements of $\mathbf{V}$. The same does not hold for a restriction to admissible aggregation functions. It is unclear how it could be established independently of clarity on the constitutive relations that a phenomenon is aggregated from its constituents in $\mathbf{V}$ by a certain type of (non-reversible) function. What type of function regulates the interplay between phenomena and constituents can only be determined *after* the constituents have been identified. Identifying the constituents, however, is exactly the purpose of Gebharter's procedure. Hence, avoiding CFC violations resulting from top-down determination by restricting the applicability of the procedure to systems with certain types of aggregation functions would render the procedure circular. Justifying the assumptions of the procedure and, thus, justifiably applying it seems to presuppose clarity on the very matter the procedure is designed to provide clarity for.

Nonetheless, let us assume for the sake of argument that there are types of mechanistic systems for which the nature of the aggregation function is known even in the absence of clarity on the constituents. Hence, the applicability of Gebharter's proposal could be confined to mechanisms known to have a non-reversible aggregation function. Yet, not even such a restriction would ensure compliance with CFC. To show this, we modify the voting example such that a vote also counts as validly cast ($\mathcal{W}=1$) if both hands are raised ($\mathcal{L}=1 \wedge \mathcal{R}=1$). The function regulating the relation between the phenomenon $\mathcal{W}$ and its constituents $\mathcal{L}$ and $\mathcal{R}$ shall hence be one of inclusive disjunction (i.e. maximum): $\mathcal{W}=1 \leftrightarrow \mathcal{L}=1 \vee \mathcal{R}=1$ (i.e. $\mathcal{W} = max(\mathcal{L}, \mathcal{R})$). While we still get bottom-up determination from this system, we no longer get top-down determination. Not every value configuration of $\{\mathcal{W}, \mathcal{R}\}$ and $\{\mathcal{W}, \mathcal{L}\}$ determines a value of $\mathcal{L}$ and $\mathcal{R}$, respectively. For example, if $\mathcal{W}=1$ and $\mathcal{L}=1$, it is not determined whether $\mathcal{R}$ takes the value 0 or 1, as both values are possible.

In order to decide whether Gebharter's procedure is reliably applicable to structures for which top-down determination can be non-circularly excluded,

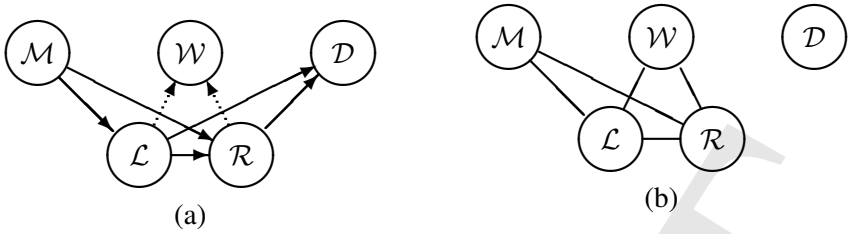

Figure 3: Voting with non-reversible aggregation function. Under epiphenomenalism*, (a) is the true graph (where dotted arrows are constitutive). In the skeleton graph (b), all edges that can be screened off in $\mathbf{O}$ have been eliminated.

we embed this non-reversible voting mechanism in a simple causal context. Let $\mathcal{M}$ be a variable representing the cause of hand raising in the voter's motor cortex, and let $\mathcal{D}$ represent the ultimate decision being taken by the vote. Let us moreover grant Gebharter that epiphenomenalism* is true. Against that background, the true structure over the variable set $\mathbf{O} = \{\mathcal{M}, \mathcal{L}, \mathcal{R}, \mathcal{W}, \mathcal{D}\}$ is given in the graph of Figure 3a. Contrary to constitutive arrows, causal arrows shall again be pseudoindeterministic, *viz.* error-disturbed, meaning that the motor cortex does not perfectly control the hand raisings, which, in turn, are not perfectly reflected in the resulting decision (e.g. due to miscounting). In that system, $\mathcal{L}$ and $\mathcal{R}$ cannot be screened off from their cause $\mathcal{M}$ by the other variables in $\mathbf{O}$. However, since $\mathcal{W}$ is a deterministic function of $\mathcal{L}$ and $\mathcal{R}$, and $\mathcal{D}$ can be expressed as a probabilistic function of $\mathcal{W}$, $\mathcal{W}$ contains all the relevant information for $\mathcal{D}$. All that matters for the decision is whether at least one hand was raised; whether it was the left or the right is irrelevant. Hence, given the value of $\mathcal{W}$ additional information about $\mathcal{L}$ or $\mathcal{R}$ has no bearing on the probability of $\mathcal{D}$. Or formally, $\mathcal{D} \perp\!\!\!\perp \mathcal{L}, \mathcal{R} \,|\, \mathcal{W}$. That is, even without top-down determination, $\mathcal{W}$ screens off the hand raisings from the resulting decision, as shown in Figure 3b, which features a skeleton graph from which all screened off edges are eliminated. These additional independencies do not follow from CMC applied to the true graph and, hence, violate CFC. Hence, not even if the absence of top-down determination could somehow be non-circularly ensured would the satisfaction of CFC be guaranteed. In sum, strengthening approach (B) by adding a restriction to certain types of aggregation functions is not a feasible option.

These findings confirm the received wisdom in the BN literature that variable sets comprising phenomena and their constituents are simply beyond the scope of warranted applicability of standard BN algorithms, such as PC, which are guaranteed to work only with pseudoindeterministic data (cf. condition 3 in Spirtes et al., 2000, 351).

3.3 PCD won't save the day

Still, one may wonder whether the principle behind Gebharter's proposal could be saved by implementing it with a *non-standard* BN algorithm that works with deterministic data, too. There indeed exists a variant of PC especially designed for variable sets featuring deterministic dependencies, *viz*. PCD (Glymour, 2007). PCD aims to make causal discovery insensitive to CFC violations induced by determinism. To this end, it operates like PC with one important exception. Unlike PC, PCD does not remove an adjacency between two variables $\mathcal{V}_i$ and $\mathcal{V}_j$ if these variables can only be rendered independent by conditionalizing on a subset of variables (excluding $\mathcal{V}_i$ and $\mathcal{V}_j$) that bring the probability of $\mathcal{V}_i$ or $\mathcal{V}_j$ up to 1. That is, screen-off relations involving maximal conditional probabilities of 1 are not taken to indicate the absence of causation. PCD only removes an adjacency between $\mathcal{V}_i$ and $\mathcal{V}_j$ if these variables can be screened-off with non-maximal conditional probabilities. If the adjacency can only be screened off with maximal probability, PCD leaves it in the graph and marks it as "uncertain" with a question mark (Glymour, 2007, 236).

The first thing to note about replacing PC by PCD in Gebharter's procedure is that discovery by PCD is much less informative than by PC. More concretely, while PC exploits conditional independencies of 1 to infer to (causal) irrelevance, PCD simply abstains from drawing any inference from deterministic independencies. As a result, whereas PC interprets the fact that phenomena and—in case of reversible aggregation functions—constituents are screened off from all outside variables as evidence for the causal isolation of mechanistic systems, PCD takes those screen-off relations to indicate that the causal embedding of mechanistic systems is unclear. That means replacing PC by PCD in Gebharter's procedure would not amount to replacing an algorithm that falsely embeds mechanistic systems in their causal environment by an algorithm that correctly embeds them but by an algorithm that does not embed them at all. PCD would hence fall short of achieving Gebharter's objective to develop a BN discovery procedure that correctly uncovers causal and constitutive relations in one go.

What is worse, it is doubtful whether the assumptions required by PCD are any more justifiable when analysing mechanistic systems than the assumptions of PC—even though PCD's assumptions are clearly weaker than PC's. While applying PC requires assuming that all conditional independencies in the data, including those with probabilities 1, faithfully reflect the true graph, applying PCD only requires assuming that the conditional independencies with probabilities lower than 1 are faithful to the true graph. But the voting example with non-reversible aggregation function (max) has shown that bottom-up determination may generate non-deterministic screen-off relations that do not follow from applying CMC to the true graph. The same thing happens in our amplifier example. Since the overall gain $\mathcal{G}$ is the sum of the individual gains $\mathcal{A}$ and $\mathcal{B}$ of the amplifier's two transistors, $\mathcal{G}$ encodes all the information on $\mathcal{A}$ and $\mathcal{B}$

relevant for the probability of distortion, $\mathcal{S}$. More concretely, even though $\mathcal{S}$ is not determined by any subset of $\mathbf{G} = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}\}$, it is screened off from $\mathcal{A}$ and $\mathcal{B}$ by the conjunction of the input $\mathcal{I}$ and the overall gain $\mathcal{G}$: if we know $\mathcal{I}$ and $\mathcal{G}$, additional information on $\mathcal{A}$ or $\mathcal{B}$ has no bearing on the probability of $\mathcal{S}$, or formally, $\mathcal{S} \perp\!\!\!\perp \mathcal{A}, \mathcal{B} \,|\, \mathcal{I}, \mathcal{G}$. These conditional independencies do not depend on $\mathcal{I}$ and $\mathcal{G}$ raising the probability of $\mathcal{S}$ to 1 and, hence, should be faithful to the true graph if PCD is applied to data on our amplifier. However, if we follow Gebharter in assuming epiphenomenalism*, the true graph is the one in Figure 2a, meaning that the downstream causal work is done by $\mathcal{A}$ and $\mathcal{B}$. According to more mainstream views, both $\mathcal{G}$ and $\{\mathcal{A}, \mathcal{B}\}$ count as causes of $\mathcal{S}$. On both views, it thus follows that these conditional independencies are unfaithful to the true graph even by the faithfulness standards of PCD.

Clearly, these (non-deterministic) CFC violations do not hinge on the particularities of the voting or the amplifier example. If a set of variables $\mathbf{D}$ determines a variable $\mathcal{V}_i$, it easily happens that $\mathcal{V}_i$ encodes all the information on $\mathbf{D}$ relevant to some downstream variable $\mathcal{V}_j$. In all such cases, $\mathcal{V}_i$ renders $\mathcal{V}_j$ conditionally independent of $\mathbf{D}$, even if the corresponding conditional probabilities are below 1. Without a doubt, this is a frequent pattern in systems featuring complete constitutive sets of phenomena. According to all metaphysical views that do not deny causal efficacy to constituents, these (non-deterministic) conditional independencies violate the faithfulness standards of PCD and, thus, render the use of that algorithm unwarranted.

3.4 False positives

Recently, there have been various studies (e.g. Zhang and Spirtes 2008, Zhalama et al. 2017) investigating to what degree violations of CFC affect the actual output of PC. These studies suggest, among other things, that proper parts of PC outputs can, under certain circumstances, be reliably interpreted *despite* CFC violations—that is, even if the rest of these outputs is unreliable. A possible response to our critique might thus be to adjust the interpretation of the outputs of BN algorithms so as to avoid fallacies induced by CFC violations. More concretely, according to the standard interpretation of outputs of BN algorithms, the presence and absence of edges reflect the presence and absence of causal relations in the true graph. As standard BN algorithms use screen-off relations as evidence for the absence of causal relations, the fact that mechanistic systems frequently generate screen-off relations unfaithful to the true graphs undermines the standard interpretation of absent edges in terms of absent causation. On the face of it, however, present edges, which reflect probabilistic dependencies that cannot be screened-off, remain unaffected by CFC violations. So perhaps there is a case to be made that, when applied to mechanistic systems, PC can still be used to reliably infer to the *presence* of causal/constitutive dependence relations without incurring false positives, even if it can*not* be used to reliably infer to the *absence* of such relations, due to a severe risk of false

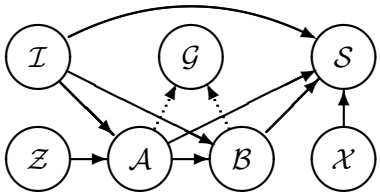

Figure 4: PC-friendly variant of the structure in Figure 2a over $\mathbf{G}^* = \mathbf{G} \cup \{\mathcal{X}, \mathcal{Z}\}$

negatives. If this holds up to scrutiny, Gebharter's approach could be used as a means to uncover the presence of constitution and causation, even if it does not reliably exhibit their absence.

To investigate that question we set up a battery of inverse search trials testing the reliability of PC's analysis of data simulated from the mechanistic structure behind our amplifier example. We conduct the trials in R using the PC implementation **pcalg** by Kalisch et al. (2012). (A replication script is available in the paper's supplementary material.) The trials have two objectives: (i) to determine the ratio of false positives both among unoriented and oriented edges issued by PC when applied to data featuring deterministic dependencies, and (ii) to determine the ratio among these false positives ascribable to the presence of determinism.

The quality of the outputs of PC (or of any other constraint-based algorithm) is known to be sensitive to various factors, for instance, to the existence of unshielded colliders, the sample size, the joint normality of the distribution or the linearity of the functional dependencies (see, e.g., Spirtes et al., 2000, 351). As deterministic dependencies induced by constitution shall be the only obstacle for PC in our trials, we ensure that they are otherwise favourable to the requirements of PC. To this end, we do not directly simulate data from the amplifier structure in Figure 2a but add two unshielded colliders, one on $\mathcal{A}$ and one on $\mathcal{S}$. More concretely, we simulate 1000 data sets with a (large) sample size of 10'000 observations each on the (epiphenomenalist*) structure in Figure 4 over the variable set $\mathbf{G}^* = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Z}\}$. We draw normally distributed values for all variables and for all (mutually independent) error terms, all being centred around 0 and having randomly sampled standard deviations. All variables are related by linear functions. To avoid that our results are sensitive to any numeric elements of those linear functions, we randomly draw numeric constants and multipliers (from the interval $[-5, 5]$) for each of the 1000 simulated data sets.

In total, we conduct two test series of the type described above, one for objective (i) and one for objective (ii). The only difference is that in the (i)-series $\mathcal{G}$ is aggregated from $\mathcal{A}$ and $\mathcal{B}$ deterministically, that is, without error terms, while in the (ii)-series $\mathcal{G}$ is a pseudoindeterministic function of $\mathcal{A}$ and $\mathcal{B}$ with an error term. All other variables, in both test series, are pseudoindeterministic.

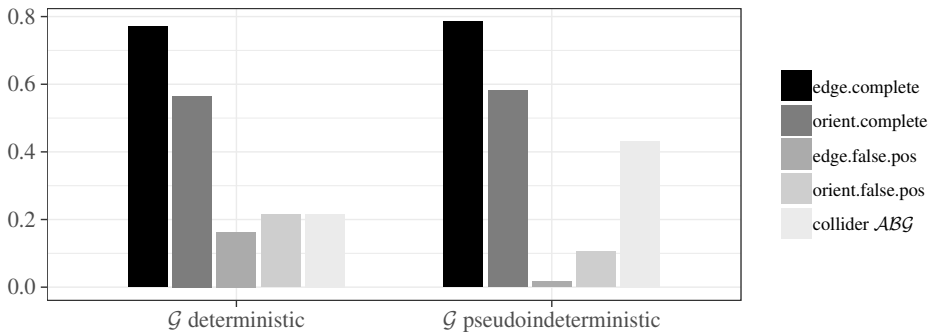

Figure 5: Completeness ratios, false positive ratios, and recovery rates of the $\mathcal{A} \longrightarrow \mathcal{G} \longleftarrow \mathcal{B}$ collider produced by (epiphenomenalist*) structures over $\mathbf{G}^*$ where the collider is deterministic (left) and pseudoindeterministic (right).

We cull false positive ratios for both unoriented edges and orientations from our tests. The false positive ratio in an individual trial is the number of unoriented/oriented edges contained in the output graph but not in the true graph of Figure 4, divided by the total number of edges in the output graph. We additionally report the completeness ratios, i.e. the number of unoriented/oriented edges contained both in the output graph and the true graph divided by the total number of edges in the true graph, as well as the recovery rate for the constitutive collider at $\mathcal{G}$, which is the core search target of Gebharter's approach. The bar chart in Figure 5 presents the means of all of the above ratios over all 1000 trials in the (i)-series on the left-hand side and the (ii)-series on the right-hand side.

Our findings show that there is a significant difference in false positive ratios. Under determinism, 16.3% of the edges output by PC are false, on average, and 21.5% of the orientations. Under pseudoindeterminism, those numbers go down to 1.6% and 10.5%, respectively. That is, $\mathcal{G}$ being a deterministic function of its constituents increases the false positive ratio for edges by a factor of 10 and for orientations by a factor of 2. Under the conditions favourable to the performance of PC, *viz.* normality, linearity, pseudoindeterminism, PC performs almost faultlessly when it comes to identifying (unoriented) edges and satisfactorily when it comes to identifying orientations. The presence of only one deterministic variable leads to 1 of 5 orientations being wrong, which is a performance hardly describable as satisfactory (under otherwise ideal discovery conditions). Importantly, the difference in false positive ratios in the two test series is not imputable to the fact that altogether fewer edges would be recovered in the pseudoindeterministic case. In fact, whether $\mathcal{G}$ is a deterministic or pseudoindeterministic function of $\mathcal{A}$ and $\mathcal{B}$ does not significantly affect the completeness ratios. In the (i)-series, on average 77% of the edges of the true graph are recovered and 56.5% of the true orientations. In the (ii)-series, those ratios go up slightly to 78.4% and 58.2%, respectively. It is also remarkable

that the recovery rate for the constitutive collider at $\mathcal{G}$ is the same under determinism as the false positive ratio for orientations. Under pseudoindeterminism, by contrast, the collider at $\mathcal{G}$ is recovered in 43.2% of the trials.

That determinism multiplies the false positive ratios of PC by 10 for edges and by 2 for orientations in trials under these—apart from determinism—ideal conditions for PC, clearly suggests a negative answer to the question whether PC could be used to reliably infer to the presence of causal/constitutive dependence relations in mechanistic systems. In our (paradigmatic) test structure, the risk of committing a false positive is the same as the chance of being rewarded by the discovery of a constitutive collider—meaning the risk is not worth taking. Gebharter could respond that a 1-in-5 false positive ratio must be tolerated if one is to make any use of BN methods for constitutive discovery, which, after all, are widely acknowledged tools for scientific modelling. In the next section, we show that this is not so.

In sum, we take the arguments in this section to cast severe doubts on Gebharter's proposed use of BN methods for constitutive discovery, and in particular, to show that treating constitution as a form of deterministic direct causation is not a promising way of bringing Bayesian methods to bear on the task of constitutive discovery. An alternative approach is required, which rejects the basic assumption that constitution is formally analogous to causation, such that BN causal discovery methods can*not* be applied to variable sets including phenomena and their constituents.

4 AN ALTERNATIVE

We cannot develop a full-blown theory of mechanistic constitution in the remainder of this paper. Instead, we confine ourselves to establishing a basis for bringing BN methods to bear on constitutive discovery in a way that avoids the aforementioned problems. To this end, we devise a sufficient condition for constitution, which, on the one hand, captures a pre-theoretic intuition many associate with constitution and, on the other, can be exploited by standard BN algorithms in a way that keeps false positive ratios low while still uncovering constitution sufficiently frequently. To be clear about an important caveat from the outset, our account—qua mere sufficient condition—will provide a handle to infer to the presence of constitution but not to its absence.

Our starting point is the view, widespread in the philosophy of the special sciences, that phenomena are causally identifiable (Fodor, 1974; Kim, 1999; Fazekas and Kertész, 2011; Soom, 2012). Here are two well-known examples from Kim (1999). Being in pain is "being in some state (or instantiating some property) caused by tissue damage and causing winces and groans" (13). Being a gene is, roughly, "the property of having some property (or being a mechanism) that performs a certain causal function, namely that of transmitting phenotypic characteristics from parents to offsprings" (10). Or, to come back to our

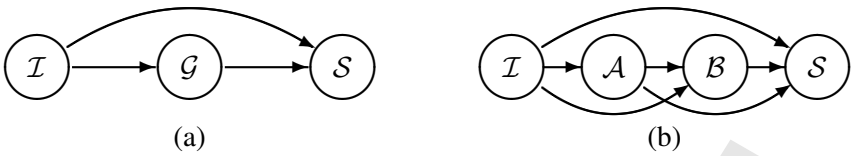

Figure 6: Causal roles (a) of $\mathcal{G}$ over $\mathbf{G} \setminus \{A, B\}$ and (b) of $\mathcal{A}$ and $\mathcal{B}$ over $\mathbf{G} \setminus \{\mathcal{G}\}$.

guiding example, amplification is that behaviour caused by voltage input and causing signal distortion. The causal identifiability of phenomena entails the falsity of epiphenomenalism: some phenomena, *viz.* the causally identifiable ones, have causes and effects.

The causal identification of a phenomenon, however, does not explain *why* that phenomenon has its characteristic causes and effects in a particular system. This is the job of a mechanistic explanation. By decomposing the phenomenon into its parts and identifying its constituents, a mechanistic explanation accounts for why the phenomenon has its characteristic causal role. Accordingly, the need for a mechanistic explanation only arises for phenomena with a characteristic causal role, that is, for phenomena with causes and effects. We do not want to impose that *all* phenomena have characteristic causal roles and, thus, can be causally identified. But clearly, if there are phenomena without causes and effects, *viz.* causally isolated phenomena, they are uninteresting from the mechanistic perspective, as they would be beyond the scope of mechanistic explainability. Overall we thus commit to a radically different metaphysical background from Gebharter's. While he endorses epiphenomenalism*, we contend that *all* mechanistically interesting phenomena have characteristic causes and effects, which is a view much more in line with the mainstream convictions in the mechanistic literature.

In a nutshell, the leading intuition underwriting our proposal is that, *if a part's causal role (partially) accounts*—in a sense to be qualified—*for why the phenomenon has its characteristic causal role, that part is a constituent*. In what follows, we make this intuition precise within the formalism of causal BNs, where the notion of the *causal role* of a variable $\mathcal{V}_i$ can be straightforwardly cashed out in terms of the set of directed edges in and out of $\mathcal{V}_i$ in the true causal BN over a variable set complying with CMC and CFC and comprising $\mathcal{V}_i$.

Our results from the previous section show that the BN machinery cannot be applied to variable sets comprising both phenomena and their parts. In consequence, before variable sets over mechanistic systems can be processed by BN methods, these variable sets must be partitioned into subsets free of mereological relations and, thereby, free of constitutive relations.[7] Contrary to $\mathbf{V}$, such constitution-free subsets (in the amplifier example, $\mathbf{G} \setminus \{\mathcal{A}, \mathcal{B}\}$ and

[7]Note that generating these mereology-free partitions presupposes (as is common in the literature on constitution) knowledge of parthood relations but not of constitutive relations.

$\mathbf{G} \setminus \{\mathcal{G}\}$; see Figure 6) can safely be assumed to comply with CMC and CFC, which makes them amenable to standard BN discovery methods.

Throughout our ensuing discussion we rely on the following formal conventions. $\mathcal{V}_1$ denotes a (mechanistically interesting) phenomenon in a variable set $\mathbf{V}$; and $\mathbf{P}_1$ denotes the set of all and only the spatiotemporal parts of $\mathcal{V}_1$ in $\mathbf{V}$—meaning that for all $\mathcal{V}_i$ in $\mathbf{P}_1$, the spatiotemporal region occupied by an instance of $\mathcal{V}_1$ contains the spatiotemporal regions occupied by the instances of $\mathcal{V}_i$. For simplicity, we assume that no other variable besides $\mathcal{V}_1$ has parts in $\mathbf{V}$—which entails that $\mathbf{P}_1$ is free of mereological relations. Moreover, $\mathbf{In}_1 \cup \mathbf{Out}_1$ denotes the set of inputs and outputs identifying $\mathcal{V}_1$'s characteristic causal role, by which we mean the causal relations between the elements of $\mathbf{In}_1 \cup \mathbf{Out}_1$ and $\mathcal{V}_1$, *viz.* the directed edges in and out of $\mathcal{V}_1$ in the true causal graph over a variable set including $\mathbf{In}_1 \cup \mathbf{Out}_1$ and $\mathcal{V}_1$ but no variables in $\mathbf{P}_1$. Since every mechanistically interesting phenomenon is causally identifiable, every such phenomenon has at least one cause and one effect. It follows that $\mathbf{In}_1 \neq \varnothing$ and $\mathbf{Out}_1 \neq \varnothing$. Finally, $\mathbf{Anc}(\mathcal{V}_i)$ and $\mathbf{Des}(\mathcal{V}_i)$ denotes the sets of, respectively, ancestors and descendants of $\mathcal{V}_i$. Then, in the true graph over $\mathbf{V} \setminus \mathbf{P}_1$, it holds that $\mathbf{In}_1 \subseteq \mathbf{Anc}(\mathcal{V}_1)$ and $\mathbf{Out}_1 \subseteq \mathbf{Des}(\mathcal{V}_1)$.

The notion of a part of $\mathcal{V}_1$ *accounting for the causal role* of $\mathcal{V}_1$, which is crucial for our leading intuition, can be spelled out in terms of that part being an element of a set of parts $\mathbf{Z}$ that has the same causal role as $\mathcal{V}_1$ and, thus, can be substituted for $\mathcal{V}_1$ in causal explanations. We shall say that $\mathbf{Z}$ has *the same causal role* as $\mathcal{V}_1$ iff (i) all variables in $\mathbf{In}_1$ have at least one effect in $\mathbf{Z}$, and (ii) all variables in $\mathbf{Out}_1$ have at least one cause in $\mathbf{Z}$, and (iii) there exists no proper subset of $\mathbf{Z}$ satisfying (i) and (ii) (i.e. $\mathbf{Z}$ is minimal with respect to (i) and (ii)). $\mathbf{Z}$ may comprise variables not contained in $\mathbf{V}$ and $\mathbf{P}_1$, respectively. Yet, even if $\mathbf{P}_1$ itself does not comprise a subset sharing $\mathcal{V}_1$'s causal role, it nonetheless holds that all variables in $\mathbf{P}_1$ located on a causal path from $\mathbf{In}_1$ to $\mathbf{Out}_1$ are contained in some such minimal set $\mathbf{Z}$ (not necessarily the same one) and, hence, account for the causal role of $\mathcal{V}_1$. It follows that all variables on a directed path from $\mathbf{In}_1$ to $\mathbf{Out}_1$ (partially) account for the causal role of $\mathcal{V}_1$ and, thus, constitute $\mathcal{V}_1$. More precisely, relative to a given variable set $\mathbf{V}$, $\mathcal{V}_1$'s causal role with respect to $\mathbf{In}_1 \cup \mathbf{Out}_1$ in $\mathbf{V} \setminus \mathbf{P}_1$ is accounted for by the parts of $\mathcal{V}_1$ on a directed path from $\mathbf{In}_1$ to $\mathbf{Out}_1$ in $\mathbf{V} \setminus \{V_1\}$. All of those parts are constituents of $\mathcal{V}_1$ in $\mathbf{V}$.

At the same time, we do not want to stipulate that all constituents account for the causal role of their phenomena. A phenomenon may have constituents that make a difference to it and yet are not contained in a minimal set sharing all of the phenomenon's causes and effects. For instance, a phenomenon may have parts causing its characteristic effects without being caused by its characteristic causes, *viz.* without being on a directed path from the latter to the former.[8] Or, our amplifier could feature parts of $\mathcal{G}$ causally influencing the

[8] We thank an anonymous referee for pointing out this possibility.

gains at $\mathcal{A}$ and $\mathcal{B}$ without being on directed paths from $\mathcal{I}$ to $\mathcal{S}$. As causes of $\mathcal{A}$ and $\mathcal{B}$, such parts would make a difference to $\mathcal{G}$ without accounting for $\mathcal{G}$'s causal role. Since we do not want to preclude the possibility that such parts are considered constituents as well, we do not elevate being on a directed path from a phenomenon's characteristic causes to its characteristic effects to the status of a necessary condition of constitution.[9]

Overall, the above considerations yield the following, causal-role (CR) based, sufficient condition for constitution:

(CR) Let $\mathcal{V}_1$'s causal role be identified by $\mathbf{In}_1 \cup \mathbf{Out}_1$, where $\mathbf{In}_1 \neq \varnothing$ and $\mathbf{Out}_1 \neq \varnothing$. Let the (true) causal graph in $\mathbf{V} \setminus \mathbf{P}_1$ be such that $\mathbf{In}_1 \subseteq \mathbf{Anc}(\mathcal{V}_1)$ and $\mathbf{Out}_1 \subseteq \mathbf{Des}(\mathcal{V}_1)$, where $\mathcal{V}_1$ is the only variable in $\mathbf{V}$ with parts in $\mathbf{V}$, and $\mathbf{P}_1$ is the set of spatiotemporal parts of $\mathcal{V}_1$ in $\mathbf{V}$. Then, $\mathcal{V}_i$ constitutes $\mathcal{V}_1$ if:

(i) $\mathcal{V}_i \in \mathbf{P}_1$; and

(ii) in the (true) causal graph over $\mathbf{V} \setminus \{\mathcal{V}_1\}$, $\mathcal{V}_i \in \mathbf{Des}(\mathbf{In}_1)$ and $\mathcal{V}_i \in \mathbf{Anc}(\mathbf{Out}_1)$.

Less formally, for a part $\mathcal{V}_i$ of a phenomenon $\mathcal{V}_1$—whose characteristic causal role is identified by the causal structure over some set $\mathbf{V} \setminus \mathbf{P}_1$—to constitute $\mathcal{V}_1$, it is sufficient that, in the causal structure over $\mathbf{V} \setminus \{\mathcal{V}_1\}$, $\mathcal{V}_i$ is on a directed path from $\mathbf{In}_1$ to $\mathbf{Out}_1$. For instance, $\mathcal{A}$ (resp. $\mathcal{B}$) constitutes $\mathcal{G}$, because there exists a variable set $\mathbf{G}$, which may be partitioned into two subsets $\mathbf{G} \setminus \mathbf{P}_\mathcal{G}$ and $\mathbf{G} \setminus \{\mathcal{G}\}$ without mereological relations, such that the structures over those subsets contain, respectively, a path $\mathcal{I} \longrightarrow \mathcal{G} \longrightarrow \mathcal{S}$, and a path $\mathcal{I} \longrightarrow \mathcal{A} \longrightarrow \mathcal{S}$ (resp. $\mathcal{I} \longrightarrow \mathcal{B} \longrightarrow \mathcal{S}$).

Our account lends itself to a straightforward methodological implementation. Given an overall set of analysed variables $\mathbf{V}$, the search target of a Bayesian discovery procedure inspired by our proposal is a set $\mathbf{C}_1$ of constituents contained in the set $\mathbf{P}_1 \subset \mathbf{V}$ of spatiotemporal parts of a target phenomenon $\mathcal{V}_1 \in \mathbf{V}$, which is causally identified by a set of characteristic causes $\mathbf{In}_1 \subset \mathbf{V}$ and effects $\mathbf{Out}_1 \subset \mathbf{V}$. According to CR, a variable $\mathcal{V}_i$ in $\mathbf{P}_1$ is contained in $\mathbf{C}_1$ if $\mathcal{V}_i$ is located on a directed path from $\mathbf{In}_1$ to $\mathbf{Out}_1$ in $\mathbf{V} \setminus \{\mathcal{V}_1\}$. To find a suitable $\mathbf{C}_1$ along these lines, $\mathbf{V}$ must first be partitioned into two distinct subsets free of mereological relations, to the effect that $\mathcal{V}_1$ and $\mathbf{P}_1$ are assigned to different partitions. Both of these partitions will be free of constitutive relations and, thus, of deterministic dependencies. It follows that they will

[9] Our proposal, thus, differs from so-called *inbetweenness* accounts of constitution, which are less cautious in that regard. They advance the condition of being a part on directed paths between the input and the output of the phenomenon as both sufficient and necessary for constitution (Harinen 2018; Prychitko 2019; cf. Craver 2015, 22). Our proposal additionally differs from such accounts insofar as the latter are formulated in terms of the notion of an intervention, whereas our aim here—like Gebharter's in his (2017b)—is to ground the inference to constitution from pure observational evidence, i.e. *without* manipulation of the mechanism.

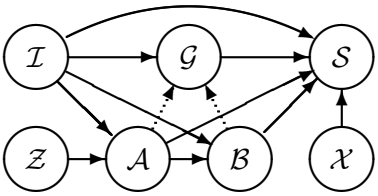

Figure 7: Non-epiphenomenalist* variant of the structure in Figure 4.

both be amenable to a standard causal analysis by BN algorithms. Assuming that $\mathcal{V}_1$ is a causally well-defined phenomenon, it follows that we know (e.g. from previous studies) that $\mathcal{V}_1$ is caused by $\mathbf{In}_1$ and causes $\mathbf{Out}_1$. That is, a causal analysis of the partition $\mathbf{V} \setminus \mathbf{P}_1$ can be used as a sort of quality benchmark for the processed data or study design. If the causal role of $\mathcal{V}_1$ is not correctly recovered, that is, if the path $\mathbf{In}_1 \longrightarrow \mathcal{V}_1 \longrightarrow \mathbf{Out}_1$ is not recovered, it can be inferred that there is a problem with the analysed data (e.g. too much noise) or with the set $\mathbf{V}$ (e.g. suitable unshielded colliders are missing), such that a causal search over $\mathbf{V} \setminus \{\mathcal{V}_1\}$ is unlikely to recover the causal roles of constituents of $\mathcal{V}_1$, either. By contrast, if this quality benchmark turns out positive, a causal analysis of the partition $\mathbf{V} \setminus \{\mathcal{V}_1\}$ is likely to identify elements of $\mathbf{C}_1$ insofar as it recovers directed causal paths from $\mathbf{In}_1$ through $\mathbf{P}_1$ into $\mathbf{Out}_1$. All parts on such paths belong to $\mathbf{C}_1$.

By rejecting the basic assumption that constitution behaves like deterministic direct causation, CR provides a simple and elegant solution to the problems incurred by Gebharter's procedure. Since CR is formulated in terms of mereology-free partitions of the total variable set $\mathbf{V}$, it is not affected by deterministic dependencies in $\mathbf{V}$ generating frequent CFC violations. This, in turn, allows for a suitable causal embedding of mechanisms both on the macro and the micro level. Likewise, a BN implementation of CR is not subject to the problem that deterministic dependencies significantly increase the false positive ratio of BN algorithms. CR-based inferences to constitution are subject to the same false positive ratios as standard causal inferences of BN algorithms.

Moreover, while Gebharter's procedure is only applicable if an analysed variable set $\mathbf{V}$ features complete constituting sets, which behave like a complete set of common causes and thus satisfy Sufficiency, CR may be correctly implemented also when an incomplete set of constituents are in $\mathbf{V}$. Provided a phenomenon's part belongs to a directed path from inputs to outputs, it counts as a constituent by the lights of CR.

To demonstrate the performance of our approach when implemented with the aid of PC, we conduct two series of inverse search trials by simulating data from the non-epiphenomenalist* version of our amplifier structure in Figure 7, which—like the structure in Figure 4—features two additional unshielded colliders to facilitate the orientation of edges. The trials are set up analogously to the trials in §3.4. (A detailed replication script is again available in the

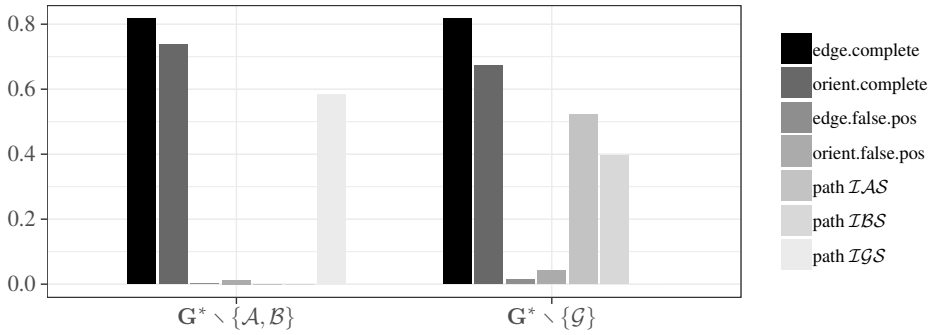

Figure 8: Completeness ratios, false positive ratios, and recovery rates for the paths the $\mathcal{I} \longrightarrow \mathcal{G} \longrightarrow \mathcal{S}$, $\mathcal{I} \longrightarrow \mathcal{A} \longrightarrow \mathcal{S}$, and $\mathcal{I} \longrightarrow \mathcal{B} \longrightarrow \mathcal{S}$ produced by the structures over $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$ (left) and over $\mathbf{G}^* \setminus \{\mathcal{G}\}$ (right).

paper's supplementary material.) In each test series, we draw 1000 data sets with 10'000 observations each; all variables in $\mathbf{G}^* = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Z}\}$ are Gaussian; all variables in $\mathbf{G}^* \setminus \{\mathcal{G}\}$ are pseudoindeterministic with mutually independent error terms; $\mathcal{G}$ is a deterministic function of $\mathcal{A}$ and $\mathcal{B}$; all functional dependencies are linear; all numeric elements of those linear functions are randomly drawn.

The first test series is run on the partition of $\mathbf{G}^*$ without the parts, *viz.* on $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$, the second series on the partition without the phenomenon $\mathbf{G}^* \setminus \{\mathcal{G}\}$. In addition to completeness and false positive ratios for both edges and orientations, we now cull the recovery rates for the directed paths from $\mathcal{I}$ via $\mathcal{G}$ or $\mathcal{A}/\mathcal{B}$ to $\mathcal{S}$ from our test results. The bar chart in Figure 8 presents the means of all of these ratios over all 1000 trials in the series over $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$ on the left-hand side and the series over $\mathbf{G}^* \setminus \{\mathcal{G}\}$ on the right-hand side.

The first and most important finding is that in both test series the false positive ratios are very low. In the partition $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$, PC produces 0.3% false edges and 1.3% false orientations, on average, while these numbers go up to 1.4% and 4.1%, respectively, in the partition $\mathbf{G}^* \setminus \{\mathcal{G}\}$.[10] Importantly, these low false positive rates are not due to PC being overly cautious in drawing inferences, as reflected by the high completeness ratios for edges (81.7% in both $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$ and $\mathbf{G}^* \setminus \{\mathcal{G}\}$) and orientations (73.8% in $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$, 67.3% in $\mathbf{G}^* \setminus \{\mathcal{G}\}$).

A second finding concerns $\mathbf{G}^* \setminus \{\mathcal{G}\}$, which, according to CR, is the partition of interest for constitutive discovery. The recovery rates for the causal paths from $\mathcal{I}$ through $\mathcal{A}$ and $\mathcal{B}$ to $\mathcal{S}$ are 52.3% and 39.5%, respectively. While those numbers are not impressive, they are significantly higher than the recovery rate (21.6%) of the constitutively revealing structural feature, *viz.* the collider at $\mathcal{G}$, Gebharter's procedure achieves in the test series in §3.4. The benchmark test

[10] Note that these results are not directly comparable with one another (or with the results in the test series of §3.4) because the true graphs relative to different variable sets differ as well.

in $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$ shows that the recovery rate (58.4%) for the macro-level path $\mathcal{I} \longrightarrow \mathcal{G} \longrightarrow \mathcal{S}$ is likewise not impressively high. This indicates that the discovery conditions for PC are not ideal in our test design. We presume that these recovery rates could be improved by, for instance, adding a further unshielded collider on $\mathcal{B}$ or another variable on the directed edge $\mathcal{I} \longrightarrow \mathcal{S}$, but we do not further investigate these variations of our test design in the present context. What matters for us here is to demonstrate the reliable applicability of CR. Whenever a PC-based implementation of CR uncovers paths from inputs of a phenomenon via its parts to its outputs, these paths can be interpreted in terms of causation at a very low false positive risk, meaning that the parts can be reliably interpreted as constituents in virtue of CR.

5 CONCLUSION

Alexander Gebharter has suggested that BN causal discovery tools may be fruitfully brought to bear on the problem of constitutive discovery. He proposes that they be used to infer to causal as well as constitutive dependencies in one go, despite the widespread view that causation and constitution are fundamentally different. The first part of this paper argued that Gebharter's proposal incurs severe problems. First, one background assumption of standard BN algorithms, *viz.* CFC, is often violated in mechanistic contexts, meaning that these algorithms—PC in particular—cannot be reliably applied. Second, the problem cannot be remedied by employing a non-standard BN algorithm, *viz.* PCD, that is designed for contexts of CFC violations induced by determinism. The reason is that PCD is much less informative and, what is worse, constitutive dependencies tend to generate probabilistic independencies that are unfaithful even by PCD's weakened Faithfulness standards. Third, only interpreting the presence (and not the absence) of edges in outputs of the PC algorithm produced in CFC-violating contexts does not amount to a promising weakening of Gebharter's proposal. We showed, in a series of inverse search trials, that determinism induced by constitution prevents PC from reliably inferring to the presence of causal/constitutive dependencies. From all this, we concluded that Gebharter's starting point, *viz.* treating constitution as a form of deterministic direct causation, and directly applying BN discovery methods to mixed sets of causal and constitutive dependencies, is not a promising way of bringing BN methods to bear on the task of constitutive discovery.

As an alternative, the second part of the paper proposed to exploit the intuition that, in a mechanistic explanation, the causal role of a target phenomenon is explained by the more fundamental causal roles of some of the system's parts. We cashed this general intuition out in the framework of BNs. More precisely, we offered a sufficient condition for constitution: if the behaviour of a part of a phenomenon is located on a directed path from the phenomenon's characteristic causes to its characteristic effects, that part is a constituent. We showed

that this condition can be tested by means of PC in a way that does not require assuming CFC (or any of the other BN assumptions for causation) to hold of variable sets including both phenomena and their parts. Our proposal avoids the problems of Gebharter's proposal in a simple and elegant way and, as a result, provides a theoretically sound foundation for the application of BN methods to constitutive discovery.

Acknowledgments We thank the audiences of Explanatory Power, Geneva, 15 June 2018, and Causation, Mechanism and Difference-Makers, Copenhagen, 3 August 2018. We are especially grateful to Alexander Gebharter, Beate Krickel, Daniel Malinsky, Alessio Moneta, and Joseph Ramsey for helpful comments and discussions. This research was generously supported by the Swiss National Science Foundation, grants no. CRSII 1_147685/1 and 100012E_160866/1 for LC and grant no. PP00P1_144736/1 for MB, and the Bergen Research Foundation, grant no. 811886 for MB.

REFERENCES

Bechtel, W. and A. Abrahamsen (2005). Explanation: a mechanist alternative. *Studies in the History and Philosophy of the Biological and Biomedical Sciences 36*, 421–41.

Craver, C. F. (2015). Levels. In T. Metzinger and J. M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group. doi: 10.15502/9783958570498.

Eronen, M. I. (2011). *Reduction in philosophy of mind: A pluralistic account*. Frankfurt am Main: Ontos.

Fazekas, P. and G. Kertész (2011). Causation at different levels: Tracking the commitments of mechanistic explanations. *Biology and Philosophy 26*, 365–83.

Fodor, J. (1974). Special sciences: Or the disunity of science as a working hypothesis. *Synthese 28*, 97–115.

Frisch, M. (2012). No place for causes? Causal skepticism in physics. *European Journal of Philosophy of Science 2*(3), 313–36.

Gebharter, A. (2017a). Causal exclusion and causal Bayes nets. *Philosophy and Phenomenological Research 95*(2), 353–75.

Gebharter, A. (2017b). Uncovering constitutive relevance relations in mechanisms. *Philosophical Studies 174*(11), 2645–66.

Gillett, C. (2002). The dimensions of realization. *Analysis 62*, 316–23.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis 44*, 49–71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science 69*(3), S342–53.

Glymour, C. (2007). Learning the structure of deterministic systems. In A. Gopnik and L. Schulz (Eds.), *Causal learning: psychology, philosophy, and computation*, pp. 231–40. Oxford University Press.

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

Kalisch, M., M. Mächler, D. Colombo, M. H. Maathuis, and P. Bühlmann (2012). Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software 47*(11), 1–26.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies 95*(1-2), 3–36.

Machamer, P., L. Darden, and C. Craver (2000). Thinking about mechanisms. *Philosophy of Science 67*, 1–25.

Norton, J. D. (2003). Causation as folk science. *Philosopher's Imprint 3*(4), 1–22.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press.

Prychitko, E. (2019). The causal situationist account of constitutive relevance. *Synthese*. doi: 10.1007/s11229-019-02170-4.

Russell, B. (1913). On the notion of cause. *Proceedings of the Aristotelian Society 13*, 1–26.

Soom, P. (2012). Mechanisms, determination and the metaphysics of neuroscience. *Studies in History and Philosophy of Biological and Biomedical Sciences 43*, 655–64.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, prediction, and search* (second ed.). Cambridge MA: MIT Press.

Spohn, W. (2006). Causation: An alternative. *The British Journal for the Philosophy of Science 57*, 93–119.

Wimsatt, W. (2007). *Re-engineering philosophy for limited beings*. Cambridge, MA: Harvard University Press.

Zhalama, J. Zhang, and W. Mayer (2017). Weakening faithfulness: some heuristic causal discovery algorithms. *International Journal of Data Science and Analytics 3*, 93–104.

Zhang, J. (2006). *Causal inference and reasoning in causally insufficient systems*. Ph. D. thesis, Department of Philosophy, Carnegie Mellon University.

Zhang, J. and P. Spirtes (2008). Detection of unfaithfulness and robust causal inference. *Minds and Machines 18*(2), 239–71.