Constitution and Causal Roles

Lorenzo Casini[†] and Michael Baumgartner[‡]

Gebharter (2017b) has proposed to use one of the best known Bayesian network (BN) causal discovery algorithms, PC, to identify the constitutive dependencies underwriting mechanistic explanations. His proposal assumes that mechanistic constitution behaves like deterministic direct causation, such that PC is directly applicable to mixed variable sets featuring both causal and constitutive dependencies. Gebharter claims that such mixed sets, under certain restrictions, comply with PC's background assumptions. 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 PC's assumptions. In the second half, we present an alternative way to bring PC to bear on the discovery of mechanistic constitution. More precisely, we argue that all of the parts of a phenomenon that 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.

The problem of constitutive discovery is: given a set of spatiotemporal parts of an explanandum phenomenon, which of these parts are explanatorily relevant, that is, constituents of the phenomenon? Importantly, clarity on parthood relations (i.e., spatiotemporal overlap) between macro and micro entities is customarily assumed by all proposed solutions of this problem (Craver, 2007; Harbecke, 2010; Couch, 2011; Gebharter, 2017b; Baumgartner and Casini, 2017; Krickel, 2018; Harinen, 2018). Parthood itself is only necessary but not sufficient for constitution and, hence, must be complemented by additional criteria in order to identify constituents. Recently, Gebharter (2017b) has suggested to bring to bear PC, one of the best known causal discovery algorithms in the causal Bayesian network (BN) framework (Spirtes et al., 2000; Pearl, 2000), on the task of identifying the constituents among a phenomenon's parts. To model and discover causation, PC identifies conditional independence constraints with statistical tests and, assuming that the analysed system satisfies certain BN axioms, causally connects variables not found to be independent. Gebharter claims that, despite the differences between causation and constitution, constitutive relations comply with the BN axioms PC assumes for causation, such that constitution can be methodologically treated as a form of (deterministic) direct causation. He concludes that PC may, together with parthood information, concurrently be applied to both causal and constitutive discovery.

After briefly introducing causal BNs (§2) and Gebharter's proposal (§3), the first part of this paper (§4) takes issue with this latter conclusion. Violations of causal BN axioms 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. Hence, these axioms may be justifiably assumed for causal contexts. By contrast, constitutive relations generate deterministic dependencies, in the presence of which violations of BN axioms—in particular, of the so-called Causal Faithfulness Condition—are no longer rare but commonplace, thus undermining their justifiable assumability. Moreover, no systematically reliable inferences can be drawn by means of PC outside the scope of validity of those axioms. We substantiate this latter point by a series of inverse search trials involving data simulations, which evaluate the performance of PC when applied to mechanistic systems.

As an alternative, the second part of the paper (§5) proposes a sufficient condition for constitution that avoids these problems and allows for bringing PC to bear on the task of constitutive discovery in a theoretically sound way.[1] In short, our proposal is that all of the parts of a phenomenon that contribute to accounting for why the phenomenon has its characteristic causal role are constituents. This idea has recently been expressed, in one way or another, by several authors (e.g., Gillett 2002, 319; see also Fazekas and Kertész 2011 and Soom 2012) but it has never been cashed out with formal precision. We fill this gap by giving it a precise rendering in the BN framework. Finally, we demonstrate the performance of our proposal by 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 $\langle \mathbf{V}, \mathbf{E}, \mathrm{Pr} \rangle$ of a finite set $\mathbf{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_\mathrm{n}\}$ 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 $\mathcal{V}_\mathrm{i}$ in the DAG obeys the Markov Condition (MC):

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

where $\mathbf{Par}_\mathrm{i}$ denotes the set of parents of $\mathcal{V}_\mathrm{i}$, and $\mathbf{Non}_\mathrm{i}$ denotes the set of non-descendants of $\mathcal{V}_\mathrm{i}$. In words, each variable is probabilistically independent of its non-descendants, conditional on its parents. Or, its parents "screen it off" from its non-descendants.

[1] Both Gebharter's proposal and our alternative to it are meant to be applied to observational data only, to avoid conceptual problems due to the inexistence of interventions in the presence of constitution (Baumgartner and Gebharter, 2016; Baumgartner and Casini, 2017). They differ from Craver's (2007) mutual manipulability theory and from so-called *inbetweenness* accounts of constitution (Craver, 2015; Harinen, 2018), all of which rely on the notion of intervention.

In a causally interpreted BN, the edges stand for direct causal relations, $\mathbf{Par}_{\mathrm{i}}$ denotes the set of $\mathcal{V}_{\mathrm{i}}$'s direct causes, $\mathbf{Non}_{\mathrm{i}}$ the set of $\mathcal{V}_{\mathrm{i}}$'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) (Spirtes et al. 2000, §3.4.1, §3.5.1).

In addition to CMC, the PC algorithm assumes the Causal Faithfulness Condition (CFC) (Zhang and Spirtes 2008, 247):

(CFC) $\langle \mathbf{V}, \mathbf{E}, \mathrm{Pr} \rangle$ 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 a probabilistic dependence—i.e., all probabilistic independencies in the graph are due to the absence of causal dependencies. In particular, CFC entails that all causal dependencies are detectable by conditional independence tests, as performed by PC.[2]

CMC and CFC are provably satisfied or only rarely violated in many well-known discovery contexts, guaranteeing that PC is reliably applicable in those contexts. On the one hand, a sufficient (albeit not necessary) condition for CMC to be provably satisfied is that (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) $\langle \mathbf{V}, \mathbf{E}, \mathrm{Pr} \rangle$ 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, a sufficient (but not necessary) condition for CFC to hold is that (i) and (ii) hold, and (v) the data contain no deterministic but only pseudoindeterministic dependencies. Then, violations of CFC have Lebesgue measure 0, entailing that they can only be produced under very strong assumptions (Spirtes et al., 2000, 42). This, in turn, is typically taken as a reason to expect them to be very rare.

At the same time, there are well-known contexts in which BN axioms are frequently violated and, hence, not justifiably assumable. One such context, relevant for the remainder of this paper, involves deterministic dependencies in the

[2] For a description of the steps of PC, the reader is referred to (Spirtes et al., 2000, 84-5).

data (which generate conditional probabilities equal to 1). Given determinism, violations of CFC are commonplace (Spirtes et al. 2000, §3.8; Glymour 2007, 236). To illustrate, whenever the dependencies along a path $\mathcal{V}_1 \longrightarrow \mathcal{V}_2 \longrightarrow \mathcal{V}_3$ are deterministic, that is, whenever $\mathcal{V}_1$ determines $\mathcal{V}_2$, which determines $\mathcal{V}_3$, it holds that $\Pr(\mathcal{V}_3 \,|\, \mathcal{V}_1 \wedge \mathcal{V}_2) = \Pr(\mathcal{V}_3 \,|\, \mathcal{V}_1) = 1$, *viz.* the indirect cause $\mathcal{V}_1$ screens off $\mathcal{V}_3$ from its direct cause $\mathcal{V}_2$. These screening-off relations, however, are not entailed by CMC, and hence violate CFC. That is, *every* deterministic chain violates CFC. The systematicity of CMC violations under determinism entails that PC is not justifiably applicable to deterministic data.

3 Gebharter's proposal

While PC is one of the most frequently discussed causal discovery tools, it has 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 forms a supervenience base and thus a sufficient condition of a phenomenon, to the effect that there cannot be change in the phenomenon without a change in its constituents. By contrast, as indicated in §2, PC is normally considered to be applicable to (pseudo)indeterministic data only.

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 absolute total voltage increase, of an amplifier subject to a voltage input $\mathcal{I}$. Amplifiers are built by assembling 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' absolute 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 also given by pseudoindeterministic functions, namely $\mathcal{A} = r_{\mathcal{A}} i - i + \epsilon_{\mathcal{A}}$ volts and $\mathcal{B} = r_{\mathcal{B}} i - i + \epsilon_{\mathcal{B}}$ volts. Assume that the first transistor amplifies by a ratio 2, and the second amplifies by a ratio 4.[3] 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 transis-

[3] This yields the amplifier's overall amplification ratio of 8 because a serial amplifier's amplification ratio of is the product of its transistors' amplification ratios.

tors' 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 axioms that PC assumes for causation. More specifically, he contends that the screening-off behaviour of complete sets of constituents (i.e., sets comprising a phenomenon's complete supervenience base) 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 are representable by causal BNs and that, with some restrictions, PC is directly applicable to variable sets featuring both constitutive and causal relations, such that the uncovered dependencies can then be grouped into causal and constitutive dependencies by using knowledge of spatiotemporal overlap (i.e., parthood relations) between instances of variables. In short, he claims that PC can perform causal and constitutive discovery in one go.

Given the well-known problems determinism creates for BN axioms, the natural conclusion to draw from Gebharter's finding that constitution behaves like deterministic direct causation would be that BNs are *incapable* of representing systems featuring constitutive relations and—*a fortiori*—PC is *inapplicable* to them. Aware that his proposal raises severe questions, Gebharter discusses two approaches to reconcile the deterministic nature of constitution with BN axioms (cf. Gebharter 2017b, 2661–62):

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

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

Approach (A) amounts to testing for determinism prior to a BN analysis (by, e.g., performing a multicollinearity test) and, if that test is positive, abstaining from applying PC. 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 behaves 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 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, too. Thus, adopting (A) in an attempt to avoid CFC violations generates frequent CMC violations, which leads Gebharter to discard (A). 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 (B), which Gebharter indeed advances as a solution to the problems prompted by the deterministic nature of constitution (Gebharter, 2017b, 2662). In §2, 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 PC be only 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* it. 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 $\mathcal{V}_3$. 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.

4 The Limits of Gebharter's Proposal

4.1 *Extensive Faithfulness violations*

Gebharter severely underestimates the problems constitutive relations induce for PC. 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 supervenience of phenomena on their constituents, every phenomenon is determined by $\mathbf{C}$. This universal *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: for any arbitrary variable $\mathcal{V}_i$, if $\mathbf{C}$ determines $\mathcal{V}_1$, then $\mathbf{C} \wedge \mathcal{V}_i$ also determines $\mathcal{V}_1$. 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 $\mathcal{V}_i$. 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), $\mathcal{G}$ (its 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 signal distortion as received by a loudspeaker. Since $\mathcal{A}$ and $\mathcal{B}$ determine $\mathcal{G}$ (from the bottom up), $\mathcal{A}$ and $\mathcal{B}$ screen off $\mathcal{G}$ from $\mathcal{I}$ and $\mathcal{S}$, or formally $\mathcal{I}, \mathcal{S} \perp\!\!\!\perp \mathcal{G} \,|\, \mathcal{A}, \mathcal{B}$.[4]

Universal bottom-up determination entails that, in CMC-warranting contexts, every mechanistic system (involving two or more levels) features condi-

[4]By contrast, $\mathcal{A}$ does not screen off $\mathcal{I}$ and $\mathcal{B}$. When holding the absolute gain of the first transistor fixed, $\mathcal{I}$ still makes a difference to the absolute gain of the second transistor. For the same reason, $\mathcal{B}$ does not screen off $\mathcal{A}$ and $\mathcal{S}$, and $\mathcal{A}$ and $\mathcal{B}$ do not screen off $\mathcal{I}$ and $\mathcal{S}$.

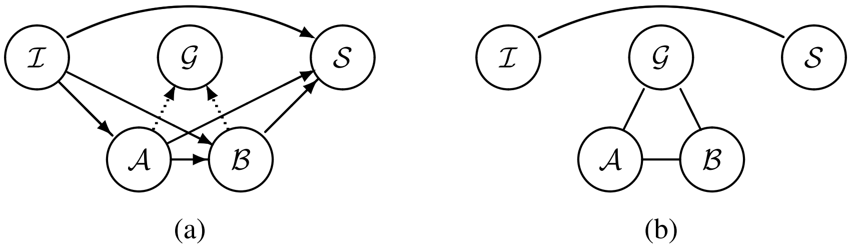

Figure 1: Graph (a) is the true structure behind a two-stage amplifier mechanism over $\mathbf{G} = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}\}$ for an epiphenomenalist*. Dotted arrows are constitutive; all other arrows are causal. Graph (b) results from applying PC to the true conditional (in)dependencies over $\mathbf{G}$, where $\mathcal{G}$ is a deterministic function of $\mathcal{A}$ and $\mathcal{B}$.

tional independencies between phenomena and all their non-constituents. Assuming CFC in such contexts implies 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, mechanists—the addressees of Gebharter's proposal—tend to be non-reductive physicalists who endorse the existence of macro-level causation.[5] They will thus reject the causal isolation of all phenomena and, consequently, refuse to assume CFC in CMC-warranting mechanistic contexts. Instead, they will interpret the independencies between phenomena and all non-constituents as yet another CFC violation induced by determinism—one that obtains even in two-level systems.

To avoid that consequence, Gebharter (2017a), in turn, rejects non-reductive physicalism and endorses a radical form of macro-level epiphenomenalism, call it *epiphenomenalism**, *viz.* the view that non-fundamental properties are 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 1a. 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. Clearly though, this manoeuvre not only clashes with the standard metaphysical commitments in the mechanistic literature but also with the scientific practice of those disciplines that are most interested in constitution, such as the social and biomedical sciences. They routinely engage in investigating causal relations among macro variables and, hence, do not commit to epiphenomenalism*.

Worse yet, 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 ab-

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

solute 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 $\mathbf{M} = \{\mathcal{G}, \mathcal{A}, \mathcal{B}\}$ is screened off from $\mathcal{I}$ and $\mathcal{S}$ by the other two elements of $\mathbf{M}$.

If PC is applied to "oracle" (true) information on conditional dependencies and independencies in $\mathbf{G}$, all edges connecting the variables in $\mathbf{M}$ to the variables in $\mathbf{G} \setminus \mathbf{M}$ will be removed, resulting in the graph skeleton in Figure 1b. This graph is *non-Markovian* because the pairs $\langle \mathcal{I}, \mathcal{A} \rangle$, $\langle \mathcal{I}, \mathcal{B} \rangle$, $\langle \mathcal{A}, \mathcal{S} \rangle$, $\langle \mathcal{B}, \mathcal{S} \rangle$ are unconnected even though these variables are pairwise unconditionally dependent. Under the assumption that CMC is satisfied, Figure 1b cannot amount to the skeleton of the true graph because too many edges have been eliminated. Moreover, since *no* Markovian graph over $\mathbf{G}$ exists that entails all the independencies depicted in Figure 1b, this constitutes a so-called *detectable* violation of CFC (Zhang and Spirtes, 2016, 252), *viz.* a CFC violation ensuing from the fact that the data cannot possibly be modelled in compliance with BN axioms.[6] No metaphysical background assumption—whether epiphenomenalism* or else—could ever reconcile the independencies in Figure 1b with CFC. The only remaining conclusion is that PC is inapplicable to our amplifier system.[7]

The possibility of top-down determination shows that not even the idiosyncratic metaphysical background of epiphenomenalism* suffices to secure the applicability of PC to mechanistic systems featuring two levels only. The crucial follow-up question now becomes how widespread top-down determination is. It is clearly not limited to amplifier gains or even to phenomena whose values are the sum of their constituents. It obtains whenever the relation between phenomena and 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

[6]The detectability of the CFC violation would render the (conservative) PC algorithm applicable *if* the CFC violation were only one of so-called *Orientation Faithfulness* and not one of *Adjacency Faithfulness* (Zhang and Spirtes, 2016, 254–55). However, what is violated here is Adjacency Faithfulness.

[7]At this point, algorithms different from PC might be resorted to, in particular, algorithms not imposing CFC (see, e.g., Malinsky and Danks 2017, 6, and Glymour et al. 2019, §4). Gebharter (2017b, 2664) himself makes some vague suggestions to the effect that PC might be replaced by other algorithms in his procedure. However, he explicitly mentions only algorithms that *weaken* CFC rather than dispense with it. Moreover, as algorithms not relying on CFC differ in informativeness and background assumptions from PC, their integration in Gebharter's procedure would require extensive procedural adjustments.

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.[8] 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), entailing 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 might be to further restrict the applicability of PC to mechanisms regulated by a *non*-reversible aggregation function, such that top-down determination does not obtain. However, such an approach would differ in a crucial respect from Gebharter's original restriction to two-level systems in (B). An analysed variable set $\mathbf{V}$ can be ensured not to feature more than two mechanistic levels by imposing that $\mathbf{V}$ does not contain a triple $\langle \mathcal{V}_i, \mathcal{V}_j, \mathcal{V}_k \rangle$ such that $\mathcal{V}_i$ is a spatiotemporal 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 identity of the constituents 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. The latter, however, is exactly the purpose of Gebharter's procedure. Hence, an attempt to avoid CFC violations resulting from top-down determination by restricting the procedure's applicability to systems with non-reversible aggregation functions would render that procedure circular.

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

[8]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 $\mathcal{W} = 1 \wedge \mathcal{R} = 1 \rightarrow \mathcal{L} = 0$.

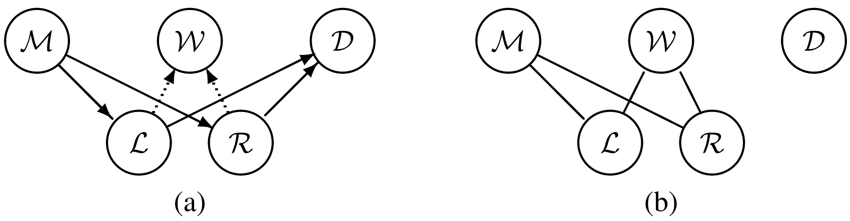

Figure 2: Voting with a non-reversible aggregation function. (a) is the true graph over **O** under epiphenomenalism* (where dotted arrows are constitutive). Graph (b) results from applying PC to the true conditional (in)dependencies over **O**.

even in the absence of clarity on the constituents. The applicability of Gebharter's proposal could thus be confined to mechanisms known to have a non-reversible aggregation function. To show that not even such a restriction would ensure compliance with CFC, 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 relation between the phenomenon $\mathcal{W}$ and its constituents $\mathcal{L}$ and $\mathcal{R}$ shall hence be regulated by the non-reversible function 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.

To decide whether Gebharter's procedure is reliably applicable to structures for which top-down determination can be non-circularly excluded, we embed this non-reversible voting mechanism in a simple causal context. Let $\mathcal{M}$ be a variable representing the cause of the hand raising in the voter's motor cortex, and let $\mathcal{D}$ represent the ultimate decision taken by the vote. Let us moreover grant Gebharter that epiphenomenalism* holds. Against that background, the true structure over $\mathbf{O} = \{\mathcal{M}, \mathcal{L}, \mathcal{R}, \mathcal{W}, \mathcal{D}\}$ is given in the graph of Figure 2a. Contrary to constitutive arrows, causal arrows shall again be pseudoindeterministic. In that system, $\mathcal{L}$ and $\mathcal{R}$ cannot be screened off from their cause $\mathcal{M}$ by the other variables in **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 $\mathcal{W}$, $\mathcal{W}$ encodes all the information on $\mathcal{L}$ and $\mathcal{R}$ relevant for the probability of $\mathcal{D}$. 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, $\mathcal{D} \perp\!\!\!\perp \mathcal{L}, \mathcal{R} \mid \mathcal{W}$. Even without top-down determination, $\mathcal{W}$ screens off the hand raisings from the resulting decision. If PC is applied to oracle information on conditional (in)dependencies in **O**, it will detach $\mathcal{D}$ from the voting mechanism, as shown in Figure 2b. Just as Figure 1b, Figure 2b is non-Markovian because the pairs $\langle \mathcal{W}, \mathcal{D} \rangle$, $\langle \mathcal{L}, \mathcal{D} \rangle$, $\langle \mathcal{R}, \mathcal{D} \rangle$ are unconnected despite the fact that they are unconditionally dependent. Since no Markovian graph exists, which is faithful to the (in)dependencies among the variables in $\mathbf{O}$, CFC is again detectably violated, which, in turn, establishes PC's inapplicability—epiphenomenalism* notwithstanding. 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 PC, which is limited to pseudoindeterministic data (cf. condition 3 in Spirtes et al., 2000, 351).

4.2 PCD won't save the day

Given the problems deterministic data generate for PC, Glymour (2007) has proposed a variant of PC, called PCD, that is custom-built for variable sets featuring deterministic dependencies. Accordingly, this section investigates whether the principle behind Gebharter's proposal could be saved by implementing it with PCD instead of PC. 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 take screen-off relations involving maximal conditional probabilities of 1 to indicate the absence of causation. PCD only infers that two variables $\mathcal{V}_i$ and $\mathcal{V}_j$ are causally unrelated if they can be screened off with non-maximal conditional probabilities. If they can only be screened off with maximal probability, the output of PCD features an edge between $\mathcal{V}_i$ and $\mathcal{V}_j$ that is marked 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. While PC exploits conditional independencies of 1 to infer to (causal) irrelevance, PCD simply abstains from drawing any inference from such independencies. Furthermore, 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 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 version of the voting example with a non-reversible aggregation function (max) has shown that bottom-up determination may generate non-deterministic screen-off relations not following from applying CMC to the true graph. The same happens in our amplifier example. Since $\mathcal{G}$ is the sum of $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{G}$ encodes all the information on $\mathcal{A}$ and $\mathcal{B}$ relevant for the probability of $\mathcal{S}$. Accordingly, although $\mathcal{S}$ 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 $\mathcal{I}$ and $\mathcal{G}$: given $\mathcal{I}$ and $\mathcal{G}$, $\mathcal{A}$ and $\mathcal{B}$ have no bearing on the probability of $\mathcal{S}$. These conditional independencies obtain despite $\mathcal{I}$ and $\mathcal{G}$ not raising the probability

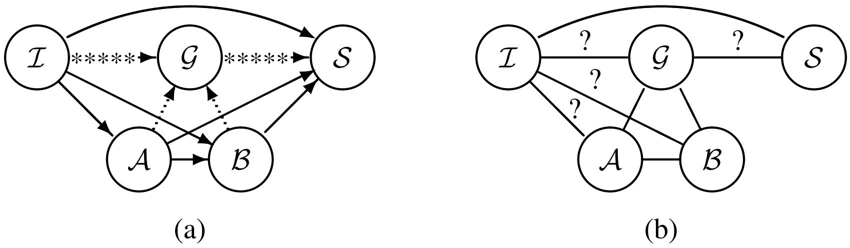

Figure 3: (a) is the true graph over $\mathbf{G}$, where starred edges mean either the absence of a dependence (epiphenomenalism*) or its presence (non-reductive physicalism). Graph (b) is the skeleton output by PCD applied to the true conditional (in)dependencies in $\mathbf{G}$, with question marks corresponding to uncertain edges.

of $\mathcal{S}$ to 1 and, hence, should be faithful to the true graph, if PCD is applied to data on our amplifier. However, they are not.

If PCD is applied to oracle information on conditional (in)dependencies among the variables in $\mathbf{G}$, its output has the skeleton in Figure 3b. Here, CFC is violated because the edges corresponding to the pairs $\langle \mathcal{A}, \mathcal{S} \rangle$ and $\langle \mathcal{B}, \mathcal{S} \rangle$ are missing, even though $\mathcal{A}$ and $\mathcal{B}$ are causes of $\mathcal{S}$ both under epiphenomenalism* and non-reductive physicalism (cf. Figure 3a). Moreover, contrary to Figures 1b and 2b, this graph is Markovian, as it preserves connections corresponding to all unconditional dependencies. In particular, the pairs $\langle \mathcal{A}, \mathcal{S} \rangle$ and $\langle \mathcal{B}, \mathcal{S} \rangle$ are connected—via $\mathcal{G}$. This means that, differently from the CFC violations incurred by PC, this CFC violation is *un*detectable.

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. Undoubtedly, this is a frequent pattern in systems featuring phenomena and complete sets of their constituents. According to all metaphysical views not denying the causal efficacy of constituents, these (non-deterministic) conditional independencies violate PCD's faithfulness standards and, thus, render the use of PCD unwarranted. Moreover, since these CFC violations are undetectable, PCD's inapplicability will tend to go unnoticed. Consequently, PCD may be unjustifiably applied resulting in *fallacious inferences*. By contrast, the detectability of the CFC violations incurred by PC ensures that PC's inapplicability does not go unnoticed, thereby preventing fallacious inferences. In sum, PCD is an even less suitable tool for constitutive inference than PC.

4.3 False positives

Recently, various studies (e.g., Zhang and Spirtes 2008, Zhalama et al. 2017) have investigated to what degree CFC violations affect the actual output of PC, among other algorithms. These studies suggest that proper parts of PC outputs

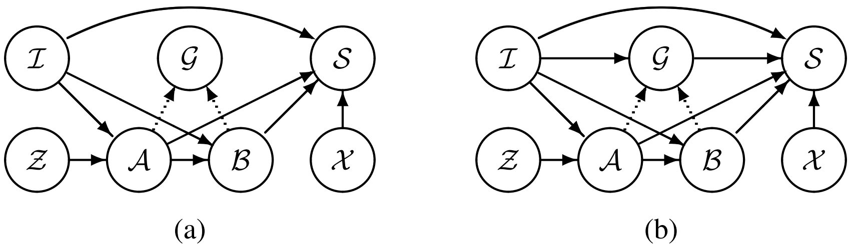

Figure 4: PC-friendly expansions of the structure in Figure 1a over $\mathbf{G}^* = \mathbf{G} \cup \{\mathcal{X}, \mathcal{Z}\}$, where (a) is the epiphenomenalist* and (b) is the non-reductive physicalist variant.

can, under certain circumstances, be reliably interpreted causally despite CFC violations. More concretely, it is CFC's purpose to ensure that the absence of edges in PC's outputs can be interpreted in terms of the absence of causation. This interpretation is blocked if CFC is violated. However, the interpretation of present edges in terms of the presence of causation remains unaffected by CFC violations. So perhaps there is a case to be made that, when applied to mechanistic systems, PC can still reliably infer the *presence* of causal/constitutive dependence relations without incurring false positives, even if it can*not* reliably infer the *absence* of such relations, due to a severe risk of false 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 false positive ratio 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 determinism.

The quality of PC's outputs is known to be sensitive to various factors, such as 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 the trials are otherwise favourable to PC. To this end, we do not directly simulate data from the amplifier structure but expand it by two unshielded colliders, one on $\mathcal{A}$ and one on $\mathcal{S}$.

The false positive ratio will, of course, depend on what we take the true data-generating structure to be. Thus, in a first test series pursuing objective (i) we grant Gebharter his epiphenomenalism* and assume that the true structure does not comprise arrows in and out of $\mathcal{G}$. The true graph over $\mathbf{G}^* = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Z}\}$ shall hence be the one in Figure 4a. We simu-

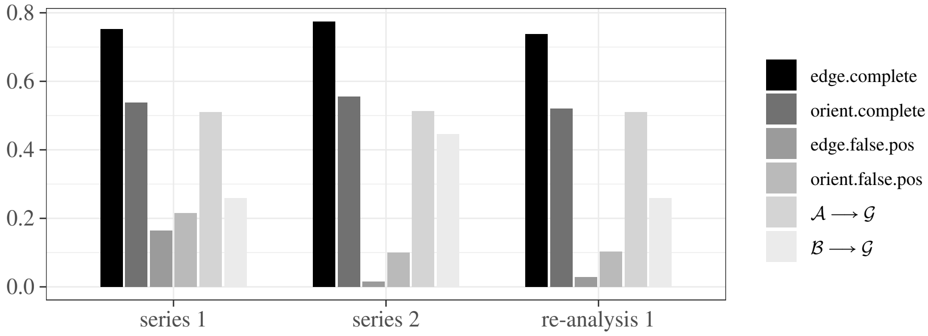

Figure 5: Completeness ratios, false positive ratios, and recovery rates for the oriented edges from $\mathcal{A}$ and $\mathcal{B}$ to $\mathcal{G}$.

late 1000 data sets with a (large) sample size of 10'000 observations from the respective data-generating structure. 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—for each of the 1000 simulated data sets—numeric constants for exogenous variables and parameters for endogenous variables (both from the interval $[-5, 5]$). In the first test series, $\mathcal{G}$ is aggregated from $\mathcal{A}$ and $\mathcal{B}$ deterministically, that is, without error terms. All other variables are pseudoindeterministic, that is, sampled with error terms.

The second test series pursues objective (ii). It differs from the first only in that $\mathcal{G}$ is a pseudoindeterministic function, too, meaning that $\mathcal{G}$ is aggregated from $\mathcal{A}$ and $\mathcal{B}$ with an error term. The true structure in the second test series is the same as that depicted in Figure 4a, with the difference that it is now interpreted as a purely causal structure in which the edges $\mathcal{A} \longrightarrow \mathcal{G}$ and $\mathcal{B} \longrightarrow \mathcal{G}$ are causal and not constitutive.

We cull false positive ratios for unoriented edges and orientations from both tests. In an individual trial, these ratios correspond to the number of unoriented/oriented edges contained in the output graph but not in the corresponding true graph of Figure 4a, divided by the total number of edges in the output graph. We additionally report completeness (or recall) ratios, that is, 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 rates for the oriented edges $\mathcal{A} \longrightarrow \mathcal{G}$ and $\mathcal{B} \longrightarrow \mathcal{G}$. The bar chart in Figure 5 shows the means of all of the above ratios over all 1000 trials in the first test series on the left-hand side and of the second series in the middle.

We find a significant difference in false positive ratios. Under determinism, on average 16.4% of the edges and 21.5% of the orientations are false. Under pseudoindeterminism, those numbers go down to 1.5% and 10.0%, respectively.

That is, $\mathcal{G}$ being a deterministic function of its constituents increases the false positive ratio for edges by a factor of 11 and for orientations by a factor of 2. Under the conditions favourable to its performance, 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 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 first test series, on average 75.2% of the edges and 53.8% of the orientations are recovered. In the second, those ratios go up slightly to 77.5% and 55.5%, respectively.

Figure 5 also shows that, under determinism, PC finds the $\mathcal{A} \longrightarrow \mathcal{G}$ connection in 51.0% of the trials but $\mathcal{B} \longrightarrow \mathcal{G}$ in only 26.0%.[9] Hence, the prospect of discovering that $\mathcal{B}$ is a constituent of $\mathcal{G}$ is almost the same as the risk of inferring a false orientation. In sum, even though PC is reasonably successful at identifying $\mathcal{A}$ as a constituent of $\mathcal{G}$ in the first series, the fact that the false positive ratio is almost as high as the recovery rate of the other constituent $\mathcal{B}$ under these—apart from determinism—ideal conditions for PC, suggests a negative answer to the question whether PC could reliably infer the presence of causal/constitutive dependence relations in mechanistic systems complying with Gebharter's background metaphysics. In our (paradigmatic) test structure, the prospect of correctly identifying the constituents of $\mathcal{G}$ is too low to counterbalance the risk of committing a false positive.

Of course, in light of our previous findings that epiphenomenalism* is insufficient to reconcile CFC with deterministic dependencies in mechanistic systems, Gebharter might renounce his endorsement of epiphenomenalism*. Accordingly, we re-analysed the data simulated in the first series against the different background assumption of non-reductive physicalism. The only difference between the first series and the re-analysis is the presupposed true graph. While we joined Gebharter in assuming $\mathcal{G}$ to be causally isolated in the first series, we now assume that $\mathcal{G}$ is on a causal path from $\mathcal{I}$ to $\mathcal{S}$, as depicted in Figure 4b.

The results are plotted in the right-hand chart of Figure 5. The numbers from the re-analysis differ from the original numbers only insofar as, upon assuming that the edges $\mathcal{I} - \mathcal{G}$ and $\mathcal{G} - \mathcal{S}$ are present in the true graph, the false positive ratio for edges and orientations drops significantly from, respectively, 16.4% and 21.5% in the epiphenomenalist* case to 2.9% and 10.4% in the non-reductive physicalist case. These results appear to suggest that switching to non-reductive physicalism is a promising move. It is, however, essential to note where this reduction of false positives comes from. In general, false edges are

[9] Under indeterminism, the recovery rates for $\mathcal{A} \longrightarrow \mathcal{G}$ are roughly the same (51.3%) but those for $\mathcal{B} \longrightarrow \mathcal{G}$ are significantly higher (44.6%).

contained in BNs if the data feature spurious dependencies or independencies that the independence tests applied by PC fail to detect. The re-analysis now shows that such a detection failure is the dominant source of the false edges in the first series. When $\mathcal{G}$ is sampled as a deterministic function of $\mathcal{A}$ and $\mathcal{B}$, these two variables raise the probability of $\mathcal{G}$ to 1, meaning they *de facto* screen $\mathcal{G}$ off from all other variables in our simulated data. Yet, those screen-off relations induced by determinism are *not systematically detected* by PC's independence test—Fisher's Z in our chosen implementation. What changes between the original analysis and the re-analysis is merely that this detection failure—which is in both cases due to the unreliability of PC's independence tests—is counted as incorrect in the former case and as correct in the latter.[10] Thus, the drop in false positive ratios under non-reductive physicalism does not establish that PC is a *reliable* tool for constitutive-and-causal inference under non-reductive physicalism but simply that PC's independence tests—presently—happen to be *unreliable* under determinism.

In sum, we take the above arguments to cast severe doubts on Gebharter's proposed use of PC 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 PC to bear on constitutive discovery. An alternative approach is required, which rejects the basic assumption that constitution is formally analogous to causation, such that PC can*not* be applied to variable sets including phenomena and their constituents.

5 An Alternative

We cannot develop a full-blown theory of mechanistic constitution in the remainder of this paper. Instead, we limit ourselves to establishing a basis for bringing PC 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 PC in a way that keeps false positive ratios low while still uncovering constitution sufficiently frequently.

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 mecha-

[10]The source of these errors is *not* the particular software implementation of PC in the **pcalg** package. Other software such as **bnlearn** (Scutari, 2010) or **Tetrad** (http://www.phil.cmu.edu/tetrad/) also fail to systematically detect these screen-off relations. Correctly detecting conditional independencies entailed by determinism using current conditional independence tests is an intricate and error-prone matter, even when the system is linear and the sample size is large.

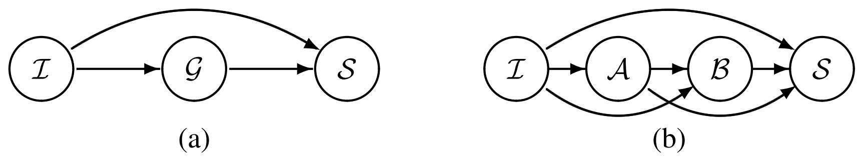

Figure 6: Causal structures (a) over $\mathbf{G} \setminus \{\mathcal{A}, \mathcal{B}\}$ and (b) over $\mathbf{G} \setminus \{\mathcal{G}\}$.

nism) that performs a certain causal function, namely that of transmitting phenotypic characteristics from parents to offsprings" (10). Or, to come back to our guiding example, amplification is that behaviour caused by voltage input and causing signal distortion. The causal identifiability of mechanistic phenomena entails a radically different metaphysical picture from epiphenomenalism*, *viz.* all mechanistic phenomena have causes and effects.

A phenomenon's causal identification, however, does not explain *why* that phenomenon has its characteristic causal role in a particular system. This is the job of a mechanistic explanation. In short, the leading intuition underwriting our proposal is that, *if a part (partially) accounts*—in a sense to be qualified—*for why the phenomenon has its characteristic causal role, that part is a constituent*. We henceforth make this intuition precise within the formalism of causal BNs over variable (sub)sets complying with CMC and CFC.

Our results from §4 show that PC should not be applied to variable sets comprising both phenomena and their parts. Before variable sets over mechanistic systems can be processed by PC, they must be partitioned into subsets free of mereological relations and, thereby, free of constitutive relations.[11] Contrary to the variable set over the whole system, such constitution-free subsets (in the amplifier example, $\mathbf{G} \setminus \{\mathcal{A}, \mathcal{B}\}$ and $\mathbf{G} \setminus \{\mathcal{G}\}$; see Figure 6) can safely be assumed to comply with CMC and CFC, which, in turn, makes them PC-processable.

Throughout our ensuing discussion we rely on the following conventions. $\mathcal{V}_1$ denotes a 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 $\mathbf{V}$—entailing 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 phenomenon is (assumed to be) causally identifiable, every 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 $\mathbf{Des}(\mathcal{V}_i)$ denote the sets

[11] Generating these mereology-free partitions presupposes knowledge of parthood relations but *not* of constitutive relations.

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 $\mathbf{Out}_1 \subseteq \mathbf{Des}(\mathcal{V}_1)$.

A necessary condition for a part of $\mathcal{V}_1$ *accounting for the characteristic causal role* of $\mathcal{V}_1$, crucial for our leading intuition, is that that part *contributes to $\mathcal{V}_1$'s role in screening off* $\mathbf{In}_1$ *and* $\mathbf{Out_1}$. That means the part belongs to a set of parts $\mathbf{Z}$, such that $\mathbf{Z}$ is substitutable for $\mathcal{V}_1$ in every $\mathbf{Sepset}(\mathbf{In}_1, \mathbf{Out_1})$ containing $\mathcal{V}_1$, that is, in every set containing $\mathcal{V}_1$ and rendering $\mathbf{In}_1$ and $\mathbf{Out}_1$ conditionally independent. Clearly, $\mathcal{V}_1$ may not suffice to screen off $\mathbf{In}_1$ and $\mathbf{Out}_1$ (cf. $\mathcal{G}$ in our amplifier example). What is important, though, is that $\mathbf{Z}$ contributes to the screening off just as much as $\mathcal{V}_1$ does.

Contributing to screen off $\mathbf{In}_1$ and $\mathbf{Out}_1$ is not sufficient for a set of parts $\mathbf{Z}$ to contain *only* constituents of $\mathcal{V}_1$. A variable set screening off $\mathbf{In}_1$ and $\mathbf{Out}_1$ may contain parts of $\mathcal{V}_1$ such that, when these parts are removed from it, the remaining variables still screen off $\mathbf{In}_1$ and $\mathbf{Out}_1$. Such parts are *redundant* for the off-screening. They do not contribute to accounting for $\mathcal{V}_1$'s role in screening off $\mathbf{In}_1$ and $\mathbf{Out}_1$ and, hence, are not constituents. In order to exclude redundant parts, $\mathbf{Z}$ must be *minimally sufficient* to play $\mathcal{V}_1$'s role in screening off $\mathbf{In}_1$ and $\mathbf{Out}_1$, meaning that no variables can be removed from $\mathbf{Z}$ such that remaining variables can still be substituted for $\mathcal{V}_1$ in every $\mathbf{Sepset}(\mathbf{In}_1, \mathbf{Out_1})$. There are two ways in which a variable can non-redundantly contribute to screen off two (sets of) other variables, namely by being a common cause of them or by being on a directed path from one to the other. Only the latter case applies to $\mathcal{V}_1$'s parts: since these parts are located temporally after $\mathbf{In}_1$ and before $\mathbf{Out}_1$, they cannot be common causes of $\mathbf{In}_1$ and $\mathbf{Out}_1$, which are prior to both. Consequently, all variables in $\mathbf{Z}$ must be on a path from $\mathbf{In}_1$ to $\mathbf{Out}_1$. Of course, $\mathbf{P}_1 \subset \mathbf{V}$ may not be rich enough to comprise a subset $\mathbf{Z}$ that plays $\mathcal{V}_1$'s role in screening off $\mathbf{In}_1$ and $\mathbf{Out}_1$. Yet, even if no such set $\mathbf{Z}$ is in $\mathbf{P}_1$, it nonetheless holds that all variables in $\mathbf{P}_1$ on a causal path from $\mathbf{In}_1$ to $\mathbf{Out}_1$ belong to some such minimal set $\mathbf{Z}$ and, hence, contribute to $\mathcal{V}_1$'s role in screening off $\mathbf{In}_1$ and $\mathbf{Out}_1$. All of those variables *constitute* $\mathcal{V}_1$.

We do not want to stipulate that all constituents account for the causal role of their phenomena. A phenomenon may have constituents making a difference to it but not contained in a minimal set $\mathbf{Z}$ sharing all of its characteristic 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. That is, our amplifier could feature parts of $\mathcal{G}$ causally influencing the gains $\mathcal{A}$ 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 characteristic causal role because they do not contribute to $\mathcal{G}$'s role in screening off $\mathcal{I}$ and $\mathcal{S}$. Since we do not want to rule out *a priori* that such parts count as constituents, 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.

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, a part $\mathcal{V}_i$ of a phenomenon $\mathcal{V}_1$—whose characteristic causal role is identified by the existence of directed paths from $\mathbf{In}_1$ to $\mathbf{Out}_1$ in the causal structure over $\mathbf{V} \setminus \mathbf{P}_1$—constitutes $\mathcal{V}_1$ if, 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, $\mathcal{A}$ and $\mathcal{B}$ constitute $\mathcal{G}$, 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 from $\mathcal{I}$ to $\mathcal{S}$ via $\mathcal{G}$, and paths from $\mathcal{I}$ to $\mathcal{S}$ via $\mathcal{A}$ and via $\mathcal{B}$ (cf. Figure 6).

Our account lends itself to a straightforward methodological implementation using PC. Given an overall set of analysed variables $\mathbf{V}$, the search target of a PC-based 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 its 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 both be PC-processable. 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 $\mathbf{Out}_1$. A causal analysis of the partition $\mathbf{V} \setminus \mathbf{P}_1$ can then be used as a sort of quality benchmark for the processed data or study design. If the paths from $\mathbf{In}_1$ to $\mathbf{Out}_1$ via $\mathcal{V}_1$ are not correctly recovered, one can infer that there is a problem with the analysed data (e.g., too much noise) or with the set $\mathbf{V}$ (e.g., missing unshielded colliders), 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 $\mathbf{V} \setminus \{\mathcal{V}_1\}$ should identify elements of $\mathbf{C}_1$ insofar as it recovers

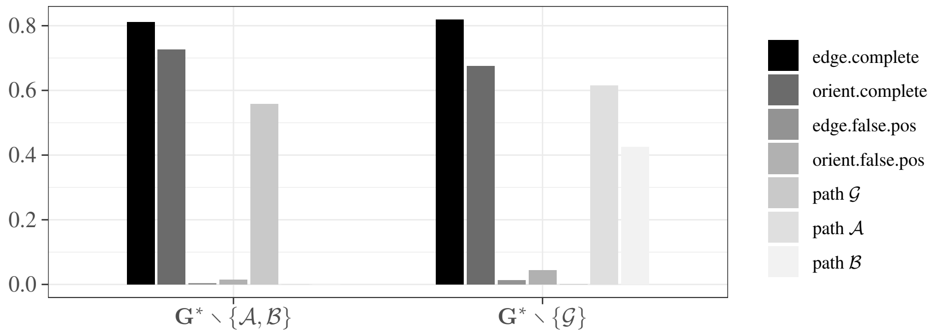

Figure 7: Completeness ratios, false positive ratios, and recovery rates for the paths via $\mathcal{G}$, $\mathcal{A}$ and $\mathcal{B}$ relative to the structures over $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$ (left) and $\mathbf{G}^* \setminus \{\mathcal{G}\}$ (right).

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 solves 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 CFC violations. This, in turn, allows for a suitable causal embedding of mechanisms both on the macro and the micro level, which, contrary to Gebharter's epiphenomenalism*, is much more in line with the mainstream convictions in the mechanistic literature. Likewise, a PC-implementation of CR is not subject to the problem that deterministic dependencies significantly increase the false positive ratio. CR-based inferences to constitution by PC are subject to the same false positive ratios as standard causal inferences by PC. Finally, 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 applied also when an incomplete set of constituents are in $\mathbf{V}$. Provided a phenomenon's part is on a directed path from the phenomenon's characteristic inputs to outputs, that part is a constituent by the lights of CR.

To demonstrate the performance of our approach when implemented with PC, we conduct another series of inverse search trials by simulating data from the non-reductive physicalist version of our amplifier structure in Figure 4b. The trials are set up analogously to those in §4.3. (A detailed replication script is again available in the paper's supplementary material.) 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.

In a first series of analyses of these 1000 data sets, PC is run on the partition of $\mathbf{G}^*$ without the parts, *viz.* on $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$, and in a second series on the partition without the phenomenon $\mathbf{G}^* \setminus \{\mathcal{G}\}$.[12] In addition to completeness and false positive ratios for both edges and orientations, we now cull the recovery rates for the directed paths from $\mathcal{I}$ to $\mathcal{S}$ via $\mathcal{G}$ and via $\mathcal{A}/\mathcal{B}$ from our test results. The bar chart in Figure 7 presents the means of all of these ratios over all 1000 trials in the series over $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$ and the series over $\mathbf{G}^* \setminus \{\mathcal{G}\}$.

The first and most important finding is that in both series the false positive ratios are very low. In the partition $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$, PC produces 0.4% false edges and 1.5% false orientations, on average, while these numbers go up to 1.4% and 4.4%, respectively, in the partition $\mathbf{G}^* \setminus \{\mathcal{G}\}$. 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 (81.2% and 82.0% in $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$ and $\mathbf{G}^* \setminus \{\mathcal{G}\}$) and orientations (72.7% in $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$, 67.6% in $\mathbf{G}^* \setminus \{\mathcal{G}\}$), nor are they due to PC's independence tests failing to spot conditional independencies induced by determinism, for the simple reason that there aren't any deterministic dependencies over $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$ and $\mathbf{G}^* \setminus \{\mathcal{G}\}$.

Secondly, the benchmark test in $\mathbf{G}^* \setminus \{\mathcal{A}, \mathcal{B}\}$ shows that the recovery rate for the macro-level path from $\mathcal{I}$ to $\mathcal{S}$ via $\mathcal{G}$, which identifies the phenomenon to be mechanistically explained, is not impressively high (55.8%). This indicates that the discovery conditions for PC are not ideal in our test design, which explains why the recovery rates for the causal paths from $\mathcal{I}$ to $\mathcal{S}$ via $\mathcal{A}$ (61.5%) and $\mathcal{B}$ (42.5%) are likewise not impressive. 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 need not investigate these variations of our test design here. What matters for us is to demonstrate the reliable applicability of CR. Whenever a PC-based implementation of CR uncovers paths from a phenomenon's inputs to its outputs via its parts, one can interpret these paths in terms of causation at a very low false positive risk, and thus reliably infer that the parts are constituents in virtue of CR.

6 Conclusion

Alexander Gebharter has suggested that the PC algorithm may be fruitfully brought to bear on the task of constitutive discovery. He proposes that it be used to infer 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 PC, *viz.* CFC, is often violated in mechanistic contexts, meaning that PC cannot be reliably applied. Second,

[12] Albeit these data are generated from the same mechanistic structure (Figure 4b) used to analyse Gebharter's procedure in §4.3, they are evaluated relative to different true graphs. Therefore, the results of the two analyses are not directly comparable.

the problem cannot be remedied by employing a modified version of PC, *viz.* PCD, that is designed for contexts of CFC violations induced by determinism. The reason is that 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 PC outputs 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 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 PC to mixed sets of causal and constitutive dependencies, is not a promising way of bringing PC 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, a phenomenon's characteristic causal role is explained by the more fundamental causal roles of some of its parts. We cashed this general intuition out in the framework of BNs. More precisely, we offered a sufficient condition for constitution: if 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 without assuming CFC of variable sets including both phenomena and their parts. Our proposal avoids the problems of Gebharter's proposal in a simple and elegant way, which provides a theoretically sound foundation for applying PC to constitutive discovery.

References

Baumgartner, M. and L. Casini (2017). An abductive theory of constitution. *Philosophy of Science 84*(2), 214–33.

Baumgartner, M. and A. Gebharter (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science 67*(3), 731–56.

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.

Couch, M. B. (2011). Mechanisms and constitutive relevance. *Synthese 183*, 375–88.

Craver, C. F. (2007). *Explaining the brain*. Oxford: Oxford University Press.

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.

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.

Glymour, C., K. Zhang, and P. Spirtes (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics 10*. article 524.

Harbecke, J. (2010). Mechanistic constitution in neurobiological explanations. *International Studies in the Philosophy of Science 24*, 267–85.

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.

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

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

Malinsky, D. and D. Danks (2017). Causal discovery algorithms: A practical guide. *Philosophy Compass*. https://doi.org/10.1111/phc3.12470.

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

Scutari, M. (2010). Learning bayesian networks with the bnlearn R package. *Journal of Statistical Software 35*(3), 1–22.

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.

Zhang, J. and P. Spirtes (2016). The three faces of faithfulness. *Synthese 193*, 1011–27.