CONSTITUTION AND CAUSAL ROLES

Lorenzo Casini[†] and Michael Baumgartner[‡]

Alexander Gebharter has recently proposed to use Bayesian network causal discovery methods to identify the constitutive dependencies that underwrite mechanistic explanations. The proposal depends on using the assumptions of the causal Bayesian network framework to implicitly define mechanistic constitution as a kind of deterministic direct causal dependence. The aim of this paper is twofold. In the first half, we argue that Gebharter's proposal incurs severe conceptual problems. In the second half, we present an alternative way to bring Bayesian network tools to bear on the issue of understanding mechanistic constitution. More precisely, our proposal interprets constitution as the relation explaining why a target phenomenon has its characteristic causal role in terms of the causal roles of some of its spatiotemporal parts—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 make a simple but paradigmatic example, 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 it causes in turn outer effects such as signal distortion, as produced by an output device (e.g., a loudspeaker). The phenomenon is explained by the augmentation of the signal by the two transistors of the amplifier arranged in series (see Wimsatt 2007, ch. 12). Another, often-cited example is the phenomenon of action potential in a neuron, namely the sudden rise and fall in the electric potential of a neuron's membrane. The phenomenon is caused by the binding of neurotransmitters, which are released by a neighbouring, "presynaptic" neuron, to receptors of the "postsynaptic"

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

neuron, and it causes in turn the release of neurotransmitters into the synaptic cleft. The phenomenon is explained by the activities of many components, most importantly by the opening and closing of voltage-gated ion channels in the neuron's membrane (see Craver 2007, chs. 2 and 4).

More generally, a mechanism is embedded in a causal context, where causal background conditions are operative that provide the external circumstances 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" (see above quote), or *constituent*, parts. The importance of the notion of constitutive relevance, or constitution, in mechanistic accounts of explanation is emphasized by Craver, who maintains that "the very idea of a mechanism presupposes the idea of constitutive relevance" (Craver, 2007, 8). While causation has been at the centre of philosophical theorizing for centuries, the notion of constitution has only recently begun to attract philosophical attention. In particular, it is still unclear what discovery method would systematize the inference to constitution.

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

In the first part of this paper, we argue that Gebharter's intended use of BNs incurs severe conceptual problems. In paradigmatic cases, allowing deterministic dependencies entails violations of a standard assumption (*viz*. Faithfulness) of BN constraint-based algorithms (in particular, PC), which Gebharter proposes to exploit for constitutional discovery, leading to systematic mistakes in causal and constitutional inference. In the second part of the paper, we propose a theory of constitution, which avoids these problems. In a nutshell, our proposal is that the constituents are all and only those parts of a phenomenon, whose causal roles account for why the phenomenon has its characteristic causal role. 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]) but it has never been cashed out in detail

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

[2] The latter, contrary to the former, maintain that the causal role of the phenomenon is identical to (rather than supervenient on) the causal role of its constituents. We do not endorse this assumption.

and with formal precision. We fill this gap by giving it a precise rendering in the framework of BNs, which is particularly suitable to explicate the notion of causal roles that figures in our account. In sum, the intended contribution of our proposal is to provide an operational definition of constitution that brings a sophisticated machinery for causal discovery, *viz.* BNs, to bear on the task of constitutional discovery in a theoretically sound way.

The paper is organized as follows. After a brief introduction to causal BNs (§2), we review Gebharter's proposal and find it flawed (§3). As an alternative, we spell out our theory, sketch its operationalization, and discuss how it avoids the discovery problems incurred by Gebharter's proposal (§4).

2 PRELIMINARIES

Let us 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. We shall use calligraphic fonts to introduce *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. This relativization to specific systems and their parts entails that mechanistic phenomena are not multiply realized by different types of systems (e.g., amplification as realized by a two-stage serial amplifier and amplification as realized by a parallel amplifier are different phenomena), albeit they may be multiply realized by multiple configurations of the values of the parts of systems of the same type (e.g., action potentials in the same neuron are overdetermined by multiple configurations of its ion channels' states, which are jointly sufficient but not individually necessary to action potential).

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 $\mathbf{E}$, whose connected nodes are the variables in $\mathbf{V}$, such that nodes and edges 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 $\mathcal{V}_i \in \mathbf{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_n\}$, $\mathcal{V}_i \perp\!\!\!\perp \mathbf{Non}_i \mid \mathbf{Par}_i$,

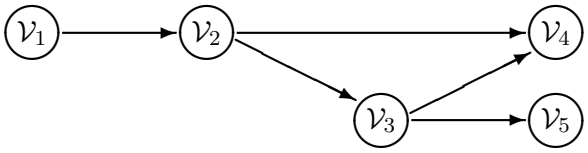


Figure 1: A Bayesian network

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 $\mathcal{V}_4 \perp\!\!\!\perp \{\mathcal{V}_1, \mathcal{V}_5\} | \{\mathcal{V}_2, \mathcal{V}_3\}$. In BN jargon, $\mathcal{V}_2$ and $\mathcal{V}_3$ "screen off" $\mathcal{V}_4$ from $\mathcal{V}_1$ and $\mathcal{V}_5$.

When the BN is causally interpreted, no pair of variables denotes properties in semantic, logical or mereological relations. In that case, the edges stand for direct causal relations, and MC is called Causal Markov Condition (CMC) (cf. Spirtes et al. 2000, §3.4.1, §3.5.1).

A precondition for a variable set to satisfy CMC is (causal) Sufficiency (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 $\mathbf{V}$ is in $\mathbf{V}$, or has the same value for all units in the population.

Sufficiency is meant to guarantee that for any two variables in $\mathbf{V}$, there is no probabilistic dependence not due to a causal dependence—that is, no probabilistic dependence is spurious. The condition entails that, while there may be latent causes of modelled variables, there are no latent *common* causes of any two modelled variables.

Finally, Faithfulness is standardly assumed in the causal BN literature (Spirtes et al. 2000, §3.4.3), especially for purposes of causal discovery:

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

Faithfulness guarantees that there is no causal dependence without probabilistic dependence—or, in other words, that the only probabilistic independencies in the graph are due to the absence of corresponding causal dependencies. For instance, the causal Faithfulness condition applied to the BN in Figure 1 implies that $\mathcal{V}_4 \not\perp\!\!\!\perp \mathcal{V}_2$, that is, there is no exact cancellation of $\mathcal{V}_2$'s effect

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

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, Faithfulness entails that all causal dependencies are detectable by conditional independence tests, as commonly performed by causal discovery algorithms (see, e.g.: Spirtes et al. 2000, 82, 88, 144; and Pearl 2000, 50, 52).

3 SHORTCOMINGS OF GEBHARTER'S PROPOSAL

While BNs have a long tradition of successful applications in causal discovery, they have played no role so far in constitutional 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 gives rise to deterministic dependencies: a complete set of constituents of a phenomenon forms 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 (but not vice versa). This sharply distinguishes constitution from causation, which in the BN framework is assumed to be non-deterministic, or "pseudoindeterministic" (cf. Spirtes et al. 2000, §3.3.2), due to latent, unmeasured causes.

To illustrate the difference, let us consider the amplification mechanism. Let $\mathcal{G}$ indicate 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. 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 a (constant) amplification ratio and $\epsilon_{\mathcal{G}}$ indicates some noise. For instance, if $\mathcal{I} = 2$ volts and the amplification ratio is 8, then $\mathcal{G} = 2 \times 8 - 2 + \epsilon_{\mathcal{G}}$ 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' absolute 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 composition function determining the overall amplification ratio of a serial amplifier is 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 overall ratio is 8. Then, when subject to an input $\mathcal{I} = 2$ volts, the first transistor amplifies the signal by $2 \times 2 - 2 + \epsilon_{\mathcal{A}}$ volts; and the second transistor receives that signal and amplifies it further by $4 \times (4 + \epsilon_{\mathcal{A}}) - (4 + \epsilon_{\mathcal{A}}) + \epsilon_{\mathcal{B}}$ volts.

By contrast, the relation between $\mathcal{G}$ on the one hand, and $\mathcal{A}$ and $\mathcal{B}$ on the other hand, is not pseudoindeterministic but deterministic: $\mathcal{A}$ and $\mathcal{B}$ determine $\mathcal{G}$. This means that, whatever noisy component is present in $\mathcal{G}$, this 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, $\epsilon_{\mathcal{G}} = 4\epsilon_{\mathcal{A}} + \epsilon_{\mathcal{B}}$.

The deterministic dependencies entailed by supervenience, in turn, tend to induce violations of one of the standard assumptions of BN discovery algorithms, *viz.* Faithfulness (Spirtes et al. 2000, §3.8; Glymour 2007, 236). Nonetheless, Gebharter (2017b) has recently contended that these algorithms can be directly applied, with some restrictions, to variable sets featuring both constitutional and causal relations, such that the uncovered dependencies are then grouped into causal and constitutional dependencies using information about temporal overlap. As a result, so Gebharter claims, BN algorithms can be used to perform causal and constitutional discovery in one go.

Even though Gebharter is not concerned with explicitly defining constitutive relevance, he still argues (2017b, 2652–54) that constitution is implicitly characterized by the same axioms that the BN framework assumes for causation. More specifically, he claims that the screening-off behaviour of complete sets of constituents 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 (2653). From that, he infers that constitutional relations can be represented by BNs and uncovered by standard BN algorithms, just like causal relations.

However, in light of the failure of Faithfulness in the presence of deterministic dependencies, the apparent conclusion to draw from Gebharter's finding that constitutive relations behave like deterministic causation would be that BNs are *not* capable of representing systems featuring constitutive relations (in addition to indeterministic causal relations), just as they are not capable of representing systems featuring deterministic causal relations, and that—*a fortiori*—BN algorithms are not applicable to systems featuring constitutive relations. Gebharter attempts to diffuse these worries by introducing restrictions on the applicability of his approach.

More specifically, he suggests (2017b, 2662) that violations of Faithfulness induced by deterministic constitutive relations can be avoided if the application of BN algorithms is restricted to mechanistic systems with two levels only. This amounts to using background knowledge on parthood relations between variable instances to only include parts of the phenomenon in the analysed variable set but not parts *of parts* of it. To illustrate, consider a mechanistic system with three levels, where $\mathcal{V}_1$ constitutively determines $\mathcal{V}_2$, which constitutively determines a phenomenon $\mathcal{V}_3$. In that case, it holds that $\Pr(\mathcal{V}_3 \,|\, \mathcal{V}_1 \wedge \mathcal{V}_2) = \Pr(\mathcal{V}_3 \,|\, \mathcal{V}_1) = 1$, *viz.* that the indirect constituent $\mathcal{V}_1$ screens off $\mathcal{V}_3$ from its direct constituent $\mathcal{V}_2$, which, however, is not entailed by CMC and hence violates Faithfulness. That is, mechanistic hierarchies with constitutive paths tend to violate Faithfulness. Without argument, Gebharter then assumes that these chainlike structures are the only type of mechanistic sys-

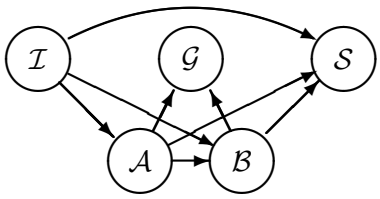


Figure 2: Structure of a two-stage amplifier mechanism, for an epiphenomenalist*.

tems that induce failures of Faithfulness due to deterministic dependencies, and recommends that BN methods be applied to two-level mechanisms only.

Yet, Gebarther severely underestimates the problems induced by his proposal. First, note that, by the assumed supervenience of phenomena on their constituents, every phenomenon has a complete set of constituents that determine it and, as a result of determination, screen it off from all other variables. It follows that a complete graph (involving any number of levels) of any mechanistic system will not feature edges in and out of phenomena. A friend of macro-level causation will interpret this finding as yet another Faithfulness violation induced by deterministic dependencies. On the assumption that phenomena *can* be caused and be causally efficacious, the screening-off relations in complete graphs of mechanistic systems exceed those entailed by CMC applied to the true graphs of these systems. What is more, this breach of Faithfulness obtains even in two-level systems, as may be illustrated by our amplification example (Figure 2). Let the variable set be $\mathbf{V}' = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}\}$, where $\mathcal{S}$ denotes a side-effect of the mechanism, namely the distortion of the signal as received by an output device, say, a loudspeaker. Given supervenience, $\{\mathcal{I}, \mathcal{S}\} \perp\!\!\!\perp \mathcal{G} \,|\, \{\mathcal{A}, \mathcal{B}\}$.

Opponents of macro-level causation can avoid this consequence. They interpret the fact that phenomena are screened off from all incoming and outgoing influences by their constituents as an argument in favour of 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. Against that background, the absence of arrows in and out of phenomena in complete mechanistic graphs is entailed by CMC and, hence, does not violate Faithfulness. Accordingly, to maintain his claim that two-level systems do not violate Faithfulness, Gebharter endorses epiphenomenalism* (cf. Gebharter, 2017a).[4] Plainly though, this consequence does not sit well with the intentions of most mechanists. While

[4]Gebharter insists (2017b, 2660) that macro variables may still be involved in "inefficient" (or "unproductive") causal relations, that is, causal relations that do not manifest themselves in difference-making patterns. Hence, Gebharter's endorses epiphenomenalism* with respect to *efficient* causation only. But his motivation for admitting inefficient relations in a formalism that takes difference making to be *necessary* for causation is unclear. For instance, it is unclear whether admitting inefficient relations delivers an advantage that outweighs their undetectability by standard BN methods and thus their indistinguishability from spurious correlations.

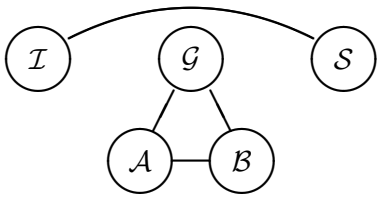


Figure 3: Graph over $\mathbf{V}' = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}\}$ inferred by PC.

there is a longstanding debate among philosophers on whether the notion of causation is dispensable in physics (Russell, 1913; Norton, 2003; Frisch, 2012), it is much less controversial to assume that causation is indispensable in other domains, such as the social and biomedical sciences, which routinely engage in testing and using causal claims. A characterization of constitution that is *a priori* incompatible with higher-level causation is at best undesirable. All the more so if macro relations pass the usual BN tests for causation in variable sets without variables in parthood relations with one another—tests the validity of which Gebharter does not dispute.

What is worse, allowing deterministic dependencies in BNs generates further problems, which—contrary to epiphenomenalism*—Gebharter cannot possibly accept. A first problem is that, in two-level systems, not only phenomena are screened off from all incoming and outgoing influences, but also constituents can be screened off in this way. To see this, reconsider the amplification example (Figure 3). $\mathcal{G}$ is an additive function of its constituents $\mathcal{A}$ and $\mathcal{B}$. Hence, not only $\mathcal{G}$ is screened off from $\mathcal{I}$ and $\mathcal{S}$ (by $\mathcal{A}$ and $\mathcal{B}$). Since $\mathcal{G}$ and $\mathcal{A}$ determine $\mathcal{B}$ (e.g., $\mathcal{G}=14 \wedge \mathcal{A}=2$ determines $\mathcal{B}=12$), also $\mathcal{B}$ is screened off from $\mathcal{I}$ and $\mathcal{S}$. Analogously, since $\mathcal{G}$ and $\mathcal{B}$ determine $\mathcal{A}$ (e.g., $\mathcal{G}=14 \wedge \mathcal{B}=12$ determines $\mathcal{A}=2$), also $\mathcal{A}$ is screened off from $\mathcal{I}$ and $\mathcal{S}$. If the graph in Figure 2 truly represents the underlying structure, these further conditional independencies are not entailed by CMC and, hence, violate Faithfulness. To avoid this consequence, Gebharter not only has to endorse the absence of causal influences in and out of $\mathcal{G}$ but also the absence of causal influences in and out of $\mathcal{A}$ and $\mathcal{B}$.

The presence of such conditional independencies is not limited to the case of additive composition functions, as may be illustrated with reference to one of Gebharter's own examples. Consider the phenomenon of bidding in an auction, $\mathcal{W}$, as constituted by raising either the left hand $\mathcal{L}$ or the right hand $\mathcal{R}$, *viz.* $\mathcal{W}=1 \leftrightarrow (\mathcal{L}=1 \wedge \mathcal{R}=0) \vee (\mathcal{L}=0 \wedge \mathcal{R}=1)$. Gebharter observes that $\mathcal{L}$ and $\mathcal{R}$ are unconditionally independent (i.e., $\mathcal{L} \perp\!\!\!\perp \mathcal{R}$), as they are caused by the bidder's free decision—call this $\mathcal{D}$—but become dependent conditional on $\mathcal{W}$ (i.e., $\mathcal{L} \not\perp\!\!\!\perp \mathcal{R} \,|\, \mathcal{W}$). From this, he concludes that "constitutive relations [...] seem to behave exactly like they were direct causal relations" (2654) in a collider structure. However, Gebharter misses an important disanalogy. If $\mathcal{L}$ and $\mathcal{R}$ *caused* (pseudoindeterministically) $\mathcal{W}$, then $\mathcal{W}$ and $\mathcal{R}$ would not

determine $\mathcal{L}$, such that $\mathcal{L} \not\perp\!\!\!\perp \mathcal{D}|\{\mathcal{W},\mathcal{R}\}$, nor would $\mathcal{W}$ and $\mathcal{L}$ determine $\mathcal{R}$, such that $\mathcal{R} \not\perp\!\!\!\perp \mathcal{D}|\{\mathcal{W},\mathcal{L}\}$. By contrast, in the constitutive case, not only $\mathcal{L}$ and $\mathcal{R}$ determine $\mathcal{W}$, such that $\mathcal{W} \perp\!\!\!\perp \mathcal{D}|\{\mathcal{L},\mathcal{R}\}$; in addition, $\mathcal{W}$ and $\mathcal{R}$ determine $\mathcal{L}$,[5] such that $\mathcal{L} \perp\!\!\!\perp \mathcal{D}|\{\mathcal{W},\mathcal{R}\}$, and $\mathcal{W}$ and $\mathcal{L}$ determine $\mathcal{R}$, such that $\mathcal{R} \perp\!\!\!\perp \mathcal{D}|\{\mathcal{W},\mathcal{L}\}$—again to the effect that not only the phenomenon but also its constituents are screened off from all non-constituents.

In sum, to maintain his claim that Faithfulness can safely be assumed for mechanistic systems with no more than two levels, Gebharter would have to contend that ever so many common mechanistic systems are causally isolated from the rest of the universe. He cannot plausibly accept this consequence, which in turn yields that the above screening-off relations must be interpreted to violate Faithfulness. All of this shows that restricting Gebharter's approach to two-level systems does not ensure that Faithfulness reliably holds for BNs featuring both causal and constitutive dependencies. In such cases, standard BN algorithms will erroneously conclude that mechanisms are isolated from their environment. In particular, Gebharter proposes to use a specific constraint-based algorithm, *viz*. PC (Spirtes et al., 2000, 84). PC infers to causation only from adjacencies, that is, from conditional dependencies. In the above cases, PC would remove all adjacencies between constituents and non-constituents, and thus find no causal relations between them (see again Figure 3).[6] That is, PC may lead to mistaken inferences when applied to mixed sets of causal and constitutive dependencies.

In response to this objection, one might qualify the proposal by maintaining that Gebharter's method is only guaranteed to work when applied to phenomena, which are multiply-realized even relative to specific systems—*viz*. phenomena, whose values are overdetermined by multiple configurations of values of their parts. Phenomena like amplification are not so overdetermined: conditional on the value of the phenomenon and all but one constituents, the value of the remaining constituent is determined. As a result, both the phenomenon and the constituents are conditionally independent of non-constituents, such that Faithfulness is violated. By contrast, when a phenomenon is overdetermined, the constituents necessarily determine the phenomenon, but for any constituent, the phenomenon and all other constituents need not determine that one constituent. As a result, the phenomenon is independent of non-constituents conditional on all constituents, in line with Gebharter's epiphenomenalism*, whereas constituents are not independent of non-constituents, conditional on the phenomenon and all other constituents, such

[5] In fact, any of the four possible value configurations of $\mathcal{W}$ and $\mathcal{L}$ suffice to fix the value of $\mathcal{R}$: $\mathcal{W}=0 \wedge \mathcal{R}=0 \rightarrow \mathcal{L}=1$; $\mathcal{W}=0 \wedge \mathcal{R}=1 \rightarrow \mathcal{L}=0$; $\mathcal{W}=1 \wedge \mathcal{R}=0 \rightarrow \mathcal{L}=1$; and $\mathcal{W}=1 \wedge \mathcal{R}=1 \rightarrow \mathcal{L}=0$. (The same point applies to $\mathcal{W}$ and $\mathcal{R}$ determining $\mathcal{L}$.)

[6] Gebharter fails to see this. He implicitly assumes that, no matter the details of the composition function relating phenomena to their constituents, PC will find dependencies between constituents and non-constituents (see 2017b, 2656, Figure 2, step 3). As our argument shows, this assumption is not always justified.

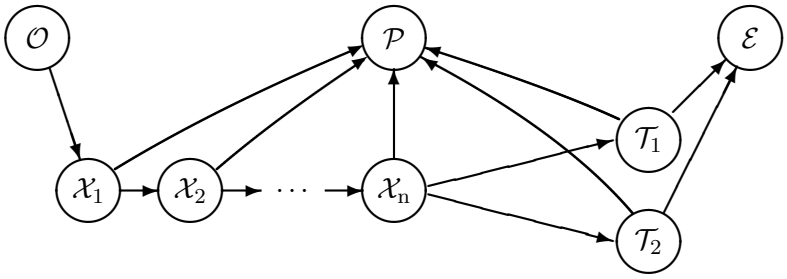


Figure 4: An epiphenomenalist* model of the action potential mechanism.

that Faithfulness is satisfied. To illustrate, one may consider a modified version of the bidding example, where raising at least one between the left hand and the right hand constitutes bidding, *viz.* $\mathcal{W} \leftrightarrow \mathcal{L} \vee \mathcal{R}$. Since $\mathcal{W}$ is overdetermined by $\mathcal{L}$ and $\mathcal{R}$, the bidder's free decision to raise a hand, $\mathcal{D}$, becomes independent of $\mathcal{W}$ conditional on $\mathcal{L}$ and $\mathcal{R}$, whereas it remains dependent of $\mathcal{L}$ (resp. $\mathcal{R}$) conditional on $\mathcal{W}$ and $\mathcal{R}$ (resp. $\mathcal{L}$), such that the $\mathcal{D}$–$\mathcal{L}$ (resp. $\mathcal{D}$–$\mathcal{R}$) adjacency is not removed. Hence, Gebharter's method would correctly analyse this revised bidding example, if its applicability were further restricted to phenomena that are overdetermined by multiple sets of constituents.

Clearly though, a thus restricted discovery procedure would hinge on a distinction between redundantly constituted phenomena and non-redundantly constituted phenomena and, in consequence, would be doomed from the start because it would presuppose knowledge *of constitution*. But not even if that circularity threat could somehow be avoided, would the restriction to overdetermined phenomena yield an adequate method, because in other cases it leads to another kind of Faithfulness violation.

To illustrate, consider the example of action potential. The mechanism works roughly as follows. Local changes in the electric potential of the neuron's membrane cause the opening/closing of ion channels, and a flow of ions through the membrane; in turn, variations in ion concentrations between the inside and the outside of the membrane determine changes in electric potential of neighbouring portions of the membrane, causing further channels to open/close; and so on. In particular, the mechanism is initiated when sodium ions flow into the neuron's body and bring the membrane's potential over a threshold, which triggers a spike along the neuron's axon. The mechanism terminates when, upon depolarization of the membranes of the neuron's terminals, calcium ions flow into the terminals and bind to vesicles filled with neurotransmitters, causing them to travel to the membrane's surface, such that neurotransmitters are released into the synaptic cleft.

Figure 4 provides a simplified model of the mechanism. Let us denote the action potential phenomenon by $\mathcal{P}$. Let us denote the cause of the mechanism—neurotransmitters binding to the membrane of the neuron's

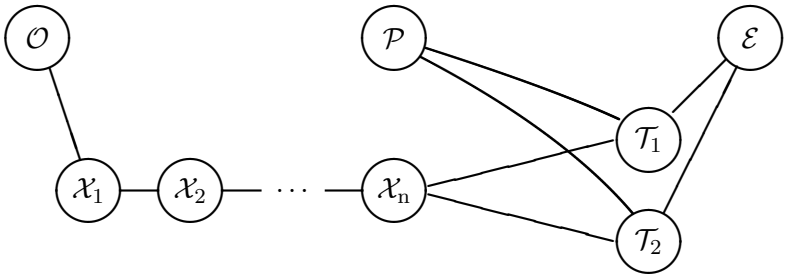


Figure 5: Graph over $\mathbf{V}'' = \{\mathcal{O}, \mathcal{P}, \mathcal{E}, \mathcal{X}_1, \ldots, \mathcal{X}_n, \mathcal{T}_1, \mathcal{T}_2\}$ inferred by PC.

body—by $\mathcal{O}$, and its effect—neurotransmitters being released by the axon terminals—by $\mathcal{E}$. Let us assume that $\mathcal{O}$ is the only exogenous cause of the mechanism. In particular, the mechanism operates against a fixed background, where subject to a suitable difference in potential, ions flow through the membrane whenever the channels open, such that ion concentrations are not independent, exogenous causes of the mechanism. To model the mechanism, let us further denote by $\mathcal{X}_i$ the opening of sodium channels in the i-th section of the axon's membrane, where i ranges from 1 (closest to the hillock) to n (closest to the terminals). To account for the fact that action potential is multiply realized, in the sense that it may obtain irrespective of whether all of the ion channels open, let us assume that, in different halves of the terminals, T_1 and T_2, neurotransmitters are released with different probabilities, because calcium channels are differently distributed on their membranes, such that different amounts of calcium ions flow into T_1 and T_2 and, as a result, different amounts of vesicles filled with neurotransmitters travel to the surface of T_1 and T_2. We use two distinct variables, $\mathcal{T}_1$ and $\mathcal{T}_2$, to represent the opening of ion channels in T_1 and T_2. Finally, let us stipulate that, as long as $\mathcal{T}_1$ *or* $\mathcal{T}_2$ are in their open state, an action potential is completed. This satisfies the further condition that the phenomenon be multiply realized, because a set of constituents including $\mathcal{T}_1$ and $\mathcal{T}_2$ redundantly determines, or overdetermines, $\mathcal{P}$.

Against this backdrop, an algorithm such as PC would infer that $\mathcal{P}$ is independent of $\mathcal{O}$ and $\mathcal{E}$ conditional on $\mathcal{T}_1$ and $\mathcal{T}_2$ (Figure 5). By contrast, $\mathcal{P}$ and $\mathcal{T}_1$ (resp. $\mathcal{T}_2$) do not determine $\mathcal{T}_2$ (resp. $\mathcal{T}_1$), such that $\mathcal{T}_1$ (resp. $\mathcal{T}_2$) is not independent of $\mathcal{O}$ and $\mathcal{E}$, in line with Gebharter's epiphenomenalism*. Here, however, the problem arises that the last two elements of the bottom-level causal path, $\mathcal{T}_1$ and $\mathcal{T}_2$, make any of their ancestors along the causal path $\mathcal{X}_1 \longrightarrow \cdots \longrightarrow \mathcal{X}_n$ redundant to determine $\mathcal{P}$, and thus to screen $\mathcal{P}$ off from $\mathcal{O}$ and $\mathcal{E}$. In other words, no state of $\mathcal{X}_1, \ldots, \mathcal{X}_n$ accounts for changes in $\mathcal{P}$ that are not already accounted for by the states of $\mathcal{T}_1$ and $\mathcal{T}_2$. Consequently, tests of conditional independence, as performed by causal discovery algorithms such as PC, will find no adjacencies between $\mathcal{P}$ and $\mathcal{X}_1, \ldots, \mathcal{X}_n$. Why is this a problem?

Theories of constitution and methods for constitutive discovery rest on a criterion for distinguishing between constitutive and non-constitutive relations. Gebharter's (2017b, 2653) characterization of constitution as a kind of direct deterministic dependence—which, to recall, was meant to avoid Faithfulness violations due to constitutive paths over more than two levels—commits him to the view that, when Sufficiency is satisfied, only constituents are adjacent to the phenomenon; any variable that is screened off from it is a non-constituent. In particular, this entails that the behaviours of parts and wholes stand in a constitutive relation only if they are adjacent. When they are not adjacent, they are not constitutively related.[7] In this respect, Gebharter's proposal (implicitly) entails that causal and constitutive dependencies are different in that variables on directed causal paths are causes of their descendants even if they are not adjacent to them, whereas variables on directed causal-or-constitutive paths are not constituents of their descendants unless they are adjacent to them. The problem with this assumption is that it is intuitively false. Some phenomena (e.g., $\mathcal{G}$) are adjacent to all of their constituents, as they would not be determined by any proper subset of them (only $\mathcal{A}$, or only $\mathcal{B}$). But not all phenomena are like this. For instance, $\mathcal{P}$ is adjacent to $\mathcal{T}_1$ and $\mathcal{T}_2$ but not to $\mathcal{X}_1, \ldots, \mathcal{X}_n$, as only the former but not the latter are necessary to determine $\mathcal{P}$. Yet, intuitively *all* such activities are constitutive of action potential, as they allow the signal to propagate from the body to the axon terminals, notwithstanding the fact that not all produce conditional dependencies, such that the distribution is unfaithful to the structure in Figure 4. To generalize, Gebharter's attempt to avoid Faithfulness violations in constitutive paths by imposing that, in two-level variable sets, constitutive dependencies are direct immediately generates further Faithfulness violations: when a phenomenon is redundantly constituted by paths of activities, some constituents are not adjacent to it. In those cases, Gebharter's proposed procedure systematically mistakes constituents for non-constituents.

In response to the aforementioned problems, one may be tempted to reply that the principle of Gebharter's proposal can be saved, provided one does not employ the traditional PC algorithm, but a version of PC that accounts for Faithfulness violations due to deterministic dependencies, *viz*. PCD, as proposed for instance by Glymour (2007).[8] In a nutshell, PCD differs from PC in that it removes an adjacency between two variables V_i and V_j only if neither one can only be made independent of the other by conditionalizing on a subset of variables (excluding V_i and V_j) that brings its probability up to 1. For instance, in the amplification example, PCD would not remove the $\mathcal{I}$–$\mathcal{A}$ adjacency, because $\mathcal{I}$ and $\mathcal{A}$ can only be made independent by conditionalizing on the subset $\{\mathcal{G}, \mathcal{B}\}$, but $\Pr(\mathcal{A} \mid \mathcal{G}, \mathcal{B}) = 1$. Analogously, in the action

[7] To wit, in Gebharter's (2017b) fictional example (Figure 1, 2650), X_4, which is a part of X_9 but not a constituent of X_9, is not adjacent to X_9.

[8] Gebharter indeed recommends using PCD in these cases (personal communication).

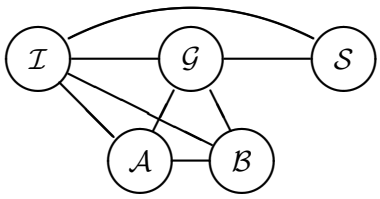


Figure 6: Graph over $\mathbf{V}' = \{\mathcal{I}, \mathcal{G}, \mathcal{S}, \mathcal{A}, \mathcal{B}\}$ inferred by PCD.

potential example, PCD would not remove the $\mathcal{X}_1$–$\mathcal{P}$ adjacency based on the conditional independence $\mathcal{X}_1 \perp\!\!\!\perp \mathcal{P} \,|\, \{\mathcal{T}_1, \mathcal{T}_2\}$, because $\Pr(\mathcal{P} \,|\, \mathcal{T}_1, \mathcal{T}_2) = 1$.

Here, however, the problem arises that other adjacencies may erroneously be removed, as may be illustrated by reference to the amplification case. Even if $\mathcal{S}$ is a (non-determininistic) function of $\mathcal{I}$, $\mathcal{A}$, and $\mathcal{B}$, it can be reduced to a function of $\mathcal{I}$ and $\mathcal{G}$ only, simply because $\mathcal{A}$ and $\mathcal{B}$ determine $\mathcal{G}$. As a consequence, it holds that $\Pr(\mathcal{S} \,|\, \mathcal{I}, \mathcal{G}) = \Pr(\mathcal{S} \,|\, \mathcal{I}, \mathcal{G}, \mathcal{A}) = \Pr(\mathcal{S} \,|\, \mathcal{I}, \mathcal{G}, \mathcal{B})$, meaning that $\mathcal{I}$ and $\mathcal{G}$ suffice to make $\mathcal{A}$ and $\mathcal{B}$ independent of $\mathcal{S}$, and bring the probability of neither $\mathcal{S}$ nor $\mathcal{A}$ or $\mathcal{B}$ up to 1. Thus, PCD will remove $\mathcal{A}$–$\mathcal{S}$ and $\mathcal{B}$–$\mathcal{S}$ (Figure 6). In contrast, $\mathcal{I}$, $\mathcal{A}$, and $\mathcal{B}$ make $\mathcal{G}$ independent of $\mathcal{S}$ but bring the probability of $\mathcal{G}$ up to 1. Thus, PCD will not remove $\mathcal{G}$–$\mathcal{S}$. In this case, not only is $\mathcal{G}$ not disconnected from the output of the mechanism, $\mathcal{S}$. What is more, the constituents $\mathcal{A}$ and $\mathcal{B}$, which according to Gebharter ought to do all the causal work, are actually made irrelevant by the phenomenon they constitute, namely $\mathcal{G}$.[9] But clearly, they are not irrelevant. By assumption, the amplifier causes distortion in virtue of the behaviours of its constituent transistors. In sum, neither PC nor PCD can deliver a sound method for causal-or-constitutive inference, as they both tend to erroneously eliminate causal or constitutive adjacencies.

We take the above arguments to undermine Gebharter's proposed use of BN algorithms for constitutive discovery, and in particular, to show that implicitly defining constitution as a form deterministic causation is not a promising way of bringing Bayesian methods to bear on the task of constitutional discovery. An alternative approach is required, which rejects the basic assumption that constitution is formally analogous to causation, such that probability distributions over variable sets including phenomena and their constituents can*not* be interpreted in the light of the BN axioms for causation, and consequently, BN causal discovery methods can*not* be applied to such variable sets.

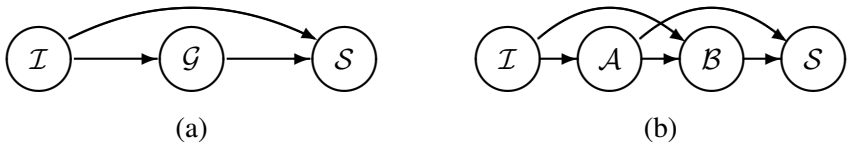
[9] This case should be distinguished from the one where epiphenomenalism* simply cannot be read off the graph, because the graph features adjacencies $\mathcal{A}$–$\mathcal{S}$ and $\mathcal{B}$–$\mathcal{S}$ *as well as* $\mathcal{G}$–$\mathcal{S}$, such that it remains undecided which dependence is "productive" and which is "unproductive" (cf. fn. 4). Here, PCD eliminates the $\mathcal{A}$–$\mathcal{S}$ and $\mathcal{B}$–$\mathcal{S}$ adjacencies, which would falsify the epiphenomenalist* assumption.

4 AN ALTERNATIVE

Epiphenomenalism is not a very popular view. Actually, in the philosophy of the special sciences, it is common to assume that phenomena are causally identified (Fodor, 1974; Kim, 1999; Fazekas and Kertész, 2011). Here are two 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). Causally identified phenomena abound also in the scientific literature. To take a textbook example, cancer is, roughly, that condition initiated by exposure to DNA-damaging factors and leading to unregulated cell growth (cf. King, 2006, 1, 24). Or, to come back to our guiding examples, amplification and action potential are role fillers in a web of other phenomena (e.g., voltage input and signal distortion in the former case, binding and release of neurotransmitters in the latter case).

Identifying the causal role of a target phenomenon with respect to other phenomena explains (ætiologically) why a given system displays that phenomenon (rather than some other phenomenon) on a particular occasion. However, it does not explain *why the phenomenon has its characteristic role* in that system. This, we take it, is the job of a mechanistic explanation. In a mechanistic explanation, the whole system (e.g., the amplifier, the neuron) instantiating the phenomenon is decomposed into its spatiotemporal parts (e.g., transistors and other elements of the circuitry, ion channels in the neuron's membrane) in order to identify the subset of parts, whose activities are constitutively responsible for the target phenomenon (e.g., amplification, action potential). What are, then, those parts? In a nutshell, our proposal is that they are all and only *those parts, whose causal roles account for why the whole has its characteristic causal role*. In this section, we formally spell out this intuition by offering a definition of constitution, which solves the problems incurred by Gebharter's proposal in a simple and elegant way.

Our proposal relativizes the notion of constitution to causally-identifiable phenomena. In particular, our proposal rejects Gebharter's epiphenomenalist* assumption, which is not assumed to follow from the causal interpretation of probability distributions over variable sets including phenomena as well as their constituents. Rather, we take at face-value the aforementioned talk of causally-identified phenomena and assume the *falsity* of epiphenomenalism*: at least some phenomena have causes and effects. Our proposal analyses the notion of constitution as relevant to such phenomena, which—we take it—are widespread in the macro-level domains of investigation where mechanistic explanations are appropriate.



(a) (b)

Figure 7: Causal roles (a) of $\mathcal{G}$ over $\mathbf{V}' \setminus \mathbf{P}_{\mathcal{G}}$ and (b) of $\mathcal{A}$ and $\mathcal{B}$ over $\mathbf{V}' \setminus \{\mathcal{G}\}$.

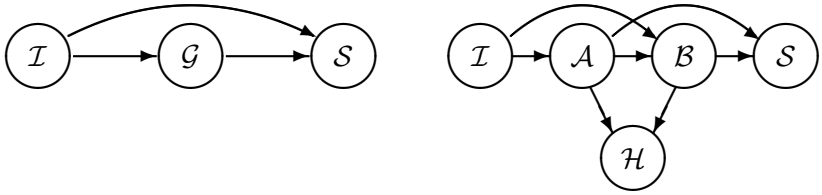
Implementing the above intuition requires—first—to identify the causal role of the phenomenon and the causal roles of its parts and—second—to identify those parts, whose causal roles account for the phenomenon's causal role. To represent causal roles, we use the formalism of causal BNs, suitably relativized to specific variables. In this context, and only in this context, we assume CMC and Faithfulness. More precisely, we do not require CMC or Faithfulness to hold in variable sets comprising both the phenomenon and its parts, such that the resulting BN is not causally interpretable, but only in subsets including one but not the other. That is, contrary to Gebharter (2017b), we exclusively adopt CMC and Faithfulness for their traditional purpose, namely *causal* inference. For us, the notion of causal roles is more basic than the notion of constitution. The latter is defined in terms of the former, but not vice versa.

Throughout our ensuing discussion we use the following definitions and make the following assumptions. $\mathcal{V}_1$ denotes a (causally-identifiable) phenomenon of interest. $\mathbf{In}_1 \cup \mathbf{Out}_1$ denotes the set of inputs and outputs that identify its characteristic causal role, by which we shall mean that the causal relations between the elements of $\mathbf{In}_1 \cup \mathbf{Out}_1$ and $\mathcal{V}_1$ are the characteristic causal role of $\mathcal{V}_1$. Since the notion of causal role presupposes the existence of at least one cause and one effect, we assume that $\mathbf{In}_1 \neq \varnothing$ and $\mathbf{Out}_1 \neq \varnothing$.

Next, $\mathbf{P}_1$ denotes the set of all and only the spatiotemporal parts of $\mathcal{V}_1$ in a variable set $\mathbf{V}$ containing $\mathcal{V}_1$, 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 $\mathcal{V}_i$. 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.

Finally, the probability distributions in $\mathbf{V} \setminus \{\mathcal{V}_1\}$ and $\mathbf{V} \setminus \mathbf{P}_1$ are causally interpretable. Let $\mathbf{Anc}(\mathcal{V}_i)$ and $\mathbf{Des}(\mathcal{V}_i)$ denote 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 $\mathbf{Out}_1 \subseteq \mathbf{Des}(\mathcal{V}_1)$. By contrast, the probability distribution over $\mathbf{V}$ is *not* causally interpretable. Hence, CMC and Faithfulness are assumed to hold in $\mathbf{V} \setminus \{\mathcal{V}_1\}$ and $\mathbf{V} \setminus \mathbf{P}_1$ but not in $\mathbf{V}$.

To illustrate the first step of our analysis, which is concerned with identifying the causal roles of phenomena and their parts, one may again consider the amplifier example. The characteristic causal role of the phenomenon $\mathcal{G}$, which consists of its causal relations to its characteristic cause, voltage input $\mathcal{I}$, and effect, signal distortion $\mathcal{S}$, may be identified in a variable set that does not include any of the amplifier's parts (Figure 7a). Similarly, the causal roles



(a) (b)

Figure 8: Causal roles (a) of $\mathcal{G}$ over $\mathbf{V}''' \setminus \mathbf{P}_{\mathcal{G}}$ and (b) of $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{H}$ over $\mathbf{V}''' \setminus \{\mathcal{G}\}$.

of $\mathcal{G}$'s parts $\mathcal{A}$ and $\mathcal{B}$ relative to $\mathcal{I}$ and $\mathcal{S}$ may be identified in a variable set that replaces $\mathcal{G}$ with its parts (Figure 7b). Notice that the first step presupposes knowledge of parthood relations but not of constitutive relations.

It should also be emphasized that this first step does not presuppose that BN algorithms can always unambiguously single out the (true) graph relative to which causal roles are identified. For instance, they will not single out the DAGs in Figure 7, and thus uncover the directed dependencies that make up the causal roles of $\mathcal{G}$ in $\mathbf{V}' \setminus \mathbf{P}_{\mathcal{G}}$ and of $\mathcal{A}$ and $\mathcal{B}$ in $\mathbf{V}' \setminus \{\mathcal{G}\}$ (due to the absence of unshielded colliders into $\mathcal{G}$ and $\mathcal{A}$). What we mean here is that—conceptually—causal roles may be represented by a BN describing the *true* causal structure over a variable set, irrespective of whether—methodologically—the discovery of that true structure is possible in *that* variable set, or some larger set is necessary (including, say, a number of unshielded colliders, whose existence we take to be guaranteed by the complexity of the world).

The second step of the analysis consists in rendering formally precise the intuition anticipated in §1: constitution is a relation among behaviours, such that the phenomenon's causal role is accounted for by its constituents' causal roles. Our proposal explicates this intuition as follows: all and only the parts of $\mathcal{V}_1$ on directed paths from $\mathbf{In}_1$ to $\mathbf{Out}_1$ in variable sets suitable for causal modelling account for $\mathcal{V}_1$'s causal role, and are thereby constituents of $\mathcal{V}_1$. 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 those of $\mathcal{V}_1$'s parts that lie on directed paths from $\mathbf{In}_1$ to $\mathbf{Out}_1$ in $\mathbf{V} \setminus \{\mathcal{V}_1\}$. All and only those parts are constituents of $\mathcal{V}_1$ in $\mathbf{V}$.

Illustrated with reference to our amplifier example, $\mathcal{G}$'s causal role in $\mathbf{V}' \setminus \mathbf{P}_{\mathcal{G}}$ is accounted for by $\mathcal{A}$'s and $\mathcal{B}$'s causal roles with respect to $\mathbf{In}_{\mathcal{G}} = \{\mathcal{I}\}$ and $\mathbf{Out}_{\mathcal{G}} = \{\mathcal{S}\}$ in $\mathbf{V}' \setminus \{\mathcal{G}\}$, because $\mathcal{A}$ and $\mathcal{B}$ are on a directed path from from $\mathcal{I}$ to $\mathcal{S}$ in $\mathbf{V}' \setminus \{\mathcal{G}\}$. Thus, $\mathcal{A}$ and $\mathcal{B}$ are the constituents of $\mathcal{G}$ in $\mathbf{V}'$. By contrast, not all parts are constituents, as may be illustrated by reference to Figure 8. Consider the variable set $\mathbf{V}''' = \mathbf{V}' \cup \{\mathcal{H}\}$, which differs from $\mathbf{V}'$ only in that it contains a variable $\mathcal{H}$, which denotes the temperature—

measured by a thermometer placed inside the amplifier—of the two transistors, such that $\mathcal{A}$ and $\mathcal{B}$ cause $\mathcal{H}$ in $\mathbf{V}'''\setminus\{\mathcal{G}\}$. $\mathcal{H}$ is what Craver would call a "sterile effect" (Craver, 2007, 143), that is, an effect of constituents that is not itself a constituent. Intuitively, $\mathcal{A}$ and $\mathcal{B}$ are *relevant* to $\mathcal{G}$'s causal role, because $\mathcal{A}$ and $\mathcal{B}$ are on a directed path from $\mathcal{I}$ to $\mathcal{S}$; by contrast, $\mathcal{H}$ is *irrelevant*, because it is *not* on a directed path from $\mathcal{I}$ to $\mathcal{S}$.

Together, the above considerations yield the following, causal-role based (CR) definition of 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, and only 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 necessary and sufficient that $\mathcal{V}_i$ is a part of $\mathcal{V}_1$, and a descendant of $\mathbf{In}_1$ and an ancestor of $\mathbf{Out}_1$ in the causal structure over $\mathbf{V} \setminus \{\mathcal{V}_1\}$—that is, that $\mathcal{V}_i$ is a part of $\mathcal{V}_1$ on a directed path from $\mathbf{In}_1$ to $\mathbf{Out}_1$.

So, for instance, $\mathcal{A}$ (resp. $\mathcal{B}$) constitutes $\mathcal{G}$, because there exists a variable set $\mathbf{V}'$, which may be partitioned into two subsets $\mathbf{V}' \setminus \mathbf{P}_\mathcal{G}$ and $\mathbf{V}' \setminus \{\mathcal{G}\}$, 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. $\mathcal{I} \longrightarrow \mathcal{B} \longrightarrow \mathcal{S}$). It is easy to verify that the definition applies also to the action potential example. In particular, any element on the causal path $\mathcal{X}_1 \longrightarrow \ldots \longrightarrow \mathcal{X}_n$ is a constituent, because there exist variable sets $\mathbf{V}'''' = \{\mathcal{O}, \mathcal{P}, \mathcal{E}, \mathcal{X}_1\}$, $\mathbf{V}''''' = \{\mathcal{O}, \mathcal{P}, \mathcal{E}, \mathcal{X}_2\}$, etc. that satisfy CR, in the sense that the required directed paths exist in the structures over $\mathbf{V}'''' \setminus \mathbf{P}_\mathcal{P}$, $\mathbf{V}'''' \setminus \{\mathcal{P}\}$, $\mathbf{V}''''' \setminus \mathbf{P}_\mathcal{P}$, $\mathbf{V}''''' \setminus \{\mathcal{P}\}$, etc.

To reiterate, not always are the correct graphs over the subsets $\mathbf{V} \setminus \mathbf{P}_1$ and $\mathbf{V} \setminus \{\mathcal{V}_1\}$ unambiguously singled out by BN algorithms (because certain unshielded colliders may be missing from $\mathbf{V}$). To that end, expansions of $\mathbf{V}$ may be necessary (containing the needed unshielded colliders). That, however, is a practical matter. Our proposal does not hinge on it: what CR requires is that directed paths exist in the true structures over $\mathbf{V} \setminus \mathbf{P}_1$ and $\mathbf{V} \setminus \{\mathcal{V}_1\}$. At the same time, relative to a variable set $\mathbf{V} \setminus \mathbf{P}_1$, where a directed path from $\mathbf{In}_1$ to $\mathbf{Out}_1$ via $\mathcal{V}_1$ exists in the causal BN inferred by BN algorithms (thanks to the existence of suitable colliders in $\mathbf{V} \setminus \mathbf{P}_1$), the required directed path from $\mathbf{In}_1$ to $\mathbf{Out}_1$ via $\mathcal{V}_i$ will also exist in the causal BN over $\mathbf{V} \setminus \{\mathcal{V}_1\}$ (thanks to the existence of corresponding colliders in $\mathbf{V} \setminus \{\mathcal{V}_1\}$).

Also, albeit CR relativizes the notion of constitution to a variable set $\mathbf{V}$, given the existence of a suitable $\mathbf{V}$, constitutive dependencies remain invariant across expansions of that $\mathbf{V}$ including, say, further parts or further causes and effects of the phenomenon. For instance, adding the behaviours of other ion channels to $\mathbf{V}''''$, such as $\mathcal{T}_1$ and $\mathcal{T}_2$, doesn't turn $\mathcal{X}_1$ into a non-constituent—irrespective of whether $\mathcal{X}_1$ is unnecessary to screen off $\mathcal{P}$ from $\mathcal{O}$ and $\mathcal{E}$, given $\mathcal{T}_1$ and $\mathcal{T}_2$.

Finally, whether a variable set $\mathbf{V}$ includes an "exhaustive set" of constituents of a phenomenon—which behave like a complete set of common causes that screen off the phenomenon from its non-constituents—is irrelevant to whether any one of them counts as a constituent. For instance, by assumption, in the amplifier's case $\mathcal{A}$ and $\mathcal{B}$ exhaustively account for $\mathcal{G}$. In $\mathbf{V}'$, both $\mathcal{A}$ and $\mathcal{B}$ are constituents of $\mathcal{G}$ and arguably form an exhaustive set of constituents. However, were $\mathcal{B}$ removed from $\mathbf{V}'$, $\mathcal{A}$ would still count as a constituent by the light of CR, even though it does not suffice to screen off $\mathcal{G}$ from $\mathcal{I}$ and $\mathcal{S}$ in $\mathbf{V}'$. The same point applies to cases where the notion of an exhaustive set is more elusive, such as the action potential case, where $\mathcal{T}_1$ and $\mathcal{T}_2$ suffice to screen off $\mathcal{P}$ from $\mathcal{O}$ and $\mathcal{E}$ in $\mathbf{V}''$, but intuitively do not form an exhaustive set of constituents. According to CR, $\mathcal{T}_1$ and $\mathcal{T}_2$ count as constituents of $\mathcal{P}$ just as much as their ancestors on the path $\mathcal{X}_1 \longrightarrow \ldots \longrightarrow \mathcal{X}_n$.

Our theory lends itself to a straightforward implementation. The search target of a Bayesian discovery procedure inspired by our proposal amounts to a set $\mathbf{C}_1$ of constituents in the subset $\mathbf{P}_1$ of spatiotemporal parts of a target phenomenon $\mathcal{V}_1$, such that $\mathbf{C}_1$ explains $\mathcal{V}_1$'s causal role. To find a suitable $\mathbf{C}_1$, one should first assign variables, whose values stand in mereological relations, to two distinct variable subsets. In particular, the subset including $\mathcal{V}_1$ should not include its parts in $\mathbf{P}_1$ and vice versa. Next, one should verify that the variables $\mathbf{In}_1$ and $\mathbf{Out}_1$ that identify the characteristic causal role of $\mathcal{V}_1$ are among, respectively, its ancestors and its descendants in $\mathbf{V} \setminus \mathbf{P}_1$. Finally, one should single out the parts that lie on directed paths from $\mathbf{In}_1$ to $\mathbf{Out}_1$ in $\mathbf{V} \setminus \{\mathcal{V}_1\}$. All and only such variables belong to $\mathbf{C}_1$, relative to $\mathbf{V}$.

By rejecting the basic assumption that constitution is (analogous to) a form of causation, our proposal provides a simple and elegant solution to the two problems incurred by Gebharter's approach. Since we do not assume Faithfulness over $\mathbf{V}$, our proposal is not subject to the problem that not all causal dependencies are identified. In particular, the identification of causal dependencies between the mechanism and its environment is guaranteed by the restriction of BN causal discovery methods to suitable subsets of $\mathbf{V}$. Nor is our proposal subject to the problem that constituents are mistaken for non-constituents when they are unnecessary to determine the phenomenon, and thus to screen it off from its non-constituents. What matters to constitution is whether, relative to a suitable subset of $\mathbf{V}$, a variable that is a spatiotemporal part of the phenomenon lies on a directed path from its causes to its effects.

Finally, we should emphasize that both our proposal and Gebharter's (2017b) proposal assume background knowledge of spatiotemporal parthood relations between values of variables. This parthood assumption is customarily made by all proposals on how to define or discover constitution in the literature (Craver, 2007; Harbecke, 2010; Couch, 2011; Gebharter, 2017b; Baumgartner and Casini, 2017; Krickel, 2018). Gebharter uses the assumption to choose variable sets with two levels of variables only; we use it to select variable sets, which may be partitioned into two subsets with variables on one level only. The sort of knowledge required to satisfy such assumptions is the same. Without doubt, ensuring that the parthood assumption is satisfied involves important epistemic as well as conceptual challenges, which make the assumption particularly "costly". What we aimed to demonstrate in this paper, though, is that anyone who is attracted by Gebharter's attempt to bring Bayesian tools to bear on the problem of constitutional discovery, should prefer our proposal to his, on the ground that it avoids its shortcomings at no additional cost. Moreover, insofar as *any* currently available theory of constitution takes parthood for granted, this point may be generalized. If one is attracted by any theory of constitution, one should be attracted by our proposal, or at least consider it a serious contender.

5 CONCLUSION

Gebharter has recently suggested that Bayesian causal discovery tools may be fruitfully brought to bear on the problem of constitutional discovery. He proposes that they be used to infer to causal as well as constitutive dependencies in one go, in spite of the widespread view that causation and constitution are different kinds of relation. The first part of this paper argued that Gebharter's proposal incurs violations of one standard assumption of BN causal discovery algorithms, *viz.* Faithfulness, leading to mistakes in causal or constitutive inference. It leads to mistakes in causal inference because, relative to certain mechanistic structures, it infers that the mechanism is isolated from its environment. And it leads to mistakes in constitutive inference because, relative to other mechanistic structures, it infers that some parts are not constituents. From this, we concluded that Gebharter's starting point, *viz.* using the assumptions of the Bayesian network formalism to implicitly define constitution as a form of direct deterministic causation, is not a promising way of bringing Bayesian networks methods to bear on the task of constitutional 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 in terms of the more fundamental causal roles of some of the system's parts. Our analysis cashed out this general intuition in the framework of Bayesian networks. More precisely, a constituent is the behaviour of a part

on a directed path from the causes to the effect that identify the characteristic causal role of the phenomenon. By not assuming Faithfulness (or any of the other BN assumptions for causation) to hold of variable sets including both phenomena and their parts, our proposal avoids in a simple and elegant way the problems of Gebharter's proposal and, as a result, provides a theoretically sound foundation to the application of BN methods to constitutional discovery.

REFERENCES

Baumgartner, M. and L. Casini (2017). An Abductive Theory of Constitution. *Philosophy of Science 84*(2), 214–33.

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.

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), 313–336.

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.

Harbecke, J. (2010). Mechanistic Constitution in Neurobiological Explanations. *International Studies in the Philosophy of Science 24*, 267–85.

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

King, R. J. B. (2006). *Cancer Biology* (3rd ed.). Pearson, Singapore.

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.

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.

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.

745 Wimsatt, W. (2007). *Re-Engineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press.

Zhang, J. (2006). *Causal Inference and Reasoning in Causally Insufficient Systems*. Ph. D. thesis, Department of Philosophy, Carnegie Mellon University. Available at http://people.hss.caltech.edu/~jiji/dissertation.pdf.