

Uncovering constitutive relevance relations in mechanisms*

Alexander Gebharter

Abstract: In this paper I argue that constitutive relevance relations in mechanisms behave like a special kind of causal relation in at least one important respect: Under suitable circumstances constitutive relevance relations produce the Markov factorization. Based on this observation one may wonder whether standard methods for causal discovery could be fruitfully applied to uncover constitutive relevance relations. This paper is intended as a first step into this new area of philosophical research. I investigate to what extent the PC algorithm, originally developed for causal search, can be used for constitutive relevance discovery. I also discuss possible objections and certain limitations of a constitutive relevance discovery procedure based on PC.

*This is a draft paper. The final version of this paper is published under the following bibliographical data: Gebharter, A. (2017). Uncovering constitutive relevance relations in mechanisms. *Philosophical Studies*, 174(11), 2645–2666. doi:10.1007/s11098-016-0803-3. The final publication is available at <http://link.springer.com>.

1 Introduction

Mechanisms play an important role in many sciences when it comes to questions concerning explanation, prediction, and control. (See, e.g., the many contributions in [Illari, Russo, & Williamson, 2011](#) explicitly referring to mechanisms.) Mechanists explain (or predict) the occurrence of certain phenomena by pointing at the mechanisms underlying these phenomena, where [Craver \(2007a, p. 6\)](#), for example, characterizes mechanisms as “entities and activities organized such that they exhibit the phenomenon to be explained. The entities are the parts. The activities are what they do.” Alternative prominent characterizations can be found in ([Machamer, Darden, & Craver, 2000](#); [Glennan, 1996](#); [Bechtel & Abrahamsen, 2005](#)). The main difference between mechanistic explanations (or predictions) and ordinary causal explanations (or predictions) is that mechanistic explanations (or predictions) not only require (i) knowledge about the causal micro structure among the system of interest’s micro parts, but also (ii) knowledge about which parts of this system are constitutively relevant for the macro behavior one wants to explain (or predict).

Nowadays causation is quite well understood and we have a multitude of sophisticated and reliable tools for achieving knowledge of type (i), i.e., for uncovering causal relationships (see, e.g., [Glymour, Spirtes, & Scheines, 1991](#); [Spirtes, Glymour, & Scheines, 2000](#); [Pearl, 2000](#)). How to achieve knowledge of type (ii), i.e., how to discover constitutive relevance relationships, on the other hand, had—at least until Craver’s (2007b) book *Explaining the brain*—rarely been investigated. Craver suggests mutual manipulability of macro and micro behavior as a criterion for uncovering constitutive relevance relationships in mechanisms. Several authors, however, have shown that Craver’s mutual manipulability approach is highly problematic (see, e.g., [Leuridan, 2012](#); [Baumgartner & Gebharder, 2016](#); [Romero, 2015](#)). The main problem seems to arise due to

the fact that it relies on ideal interventions à la [Woodward \(2003\)](#) and that it is impossible to surgically intervene on a mechanism's macro behavior, i.e., to directly causally influence a mechanism's macro behavior without directly causally influencing the behavior of some of its micro parts, simply because a mechanism's macro behavior is assumed to supervene on the behavior of this mechanism's micro parts ([Baumgartner & Gebharder, 2016](#)).

There is a number of very recent alternative suggestions of how constitutive relevance relations could be uncovered. [Baumgartner and Gebharder \(2016\)](#), for example, propose to modify Woodward's (2003) interventionist theory of causation and supplement Craver's (2007a, 2007b) mutual manipulability criterion with what they call the fat-handedness criterion. They argue that if a macro variable and a micro variable are mutually manipulable within this modified interventionist theory and, at the same time, the fat-handedness criterion is satisfied, then constitutive relevance could be abductively inferred. This idea of an abductive approach to uncover constitutive relevance relations is further developed by [Baumgartner and Casini \(in press\)](#). Another way to infer constitutive relevance relations is developed by [Harbecke \(2015\)](#). Harbecke suggests a regularity theory of constitution and constitutional inference. The theory does, however, rely on the regularity theory of causation. Unfortunately it shares all the problems that come with that particular theory.

The goal of this paper is to investigate yet another possible way to go. The basic idea underlying the account developed in this paper stems from my dissertation thesis ([Gebharder, in press](#)). It does not rely on strict regularities. It also does not rely on interventions, which might be an advantage, since how exactly interventions work in systems featuring variables standing in other than causal relations is still not well understood (cf. [Woodward, 2015](#); [Gebharder, 2015](#)). The project to be started with this paper is to some extent inspired by

Ramsey, Gazis, Roush, Spirtes, and Glymour (2002). Ramsey et al. propose to use a modified version of the PC algorithm (Spirtes et al., 2000, pp. 84f)—which was originally designed for causal discovery—to identify components (or constituents) of rock and soil samples. They found that their search procedure is able to identify the most frequently occurring kinds of carbonates equally well or even better than human experts. The key point here is that PC, a causal search algorithm, was successfully used for a kind of non-causal discovery. The basic idea underlying the present paper is that PC might also be used to uncover constitutive relevance relations in mechanisms.

This new project can be expected to encounter heavy resistance. It seems to be consensus in the philosophical literature that constitutive relations do not behave like causal relations in many crucial respects (see, e.g., Craver & Bechtel, 2007). In addition, philosophers interested in mechanisms are typically sceptical about the applicability of causal Bayes net methods to mechanisms in general, mainly because there seem to be unresolvable issues with the two most important core conditions of that framework: the causal Markov condition and the causal faithfulness condition. Within the sections of this paper, I will argue that this worry is to some extent legitimate, but—at the same time—much too pessimistic. It will turn out that there are circumstances and situations in which PC actually can be fruitfully used to uncover constitutive relevance relations in mechanisms. The search procedure for constitutive relevance to be investigated has, like any other search procedure, certain limitations.

The paper is structured as follows: I will briefly introduce some characteristic marks of constitutive relevance relations in section 2. In section 3 I will present the central notions and axioms of the causal Bayes net formalism and argue that this formalism can also be used to capture constitutive relevance. In particular, I will argue that constitutive relevance relations share one impor-

tant formal property with direct causal relations: Under certain circumstances they produce the Markov factorization. If this diagnosis is correct, then standard search procedures for causal relations should also be applicable to uncover constitutive relevance relations. In [section 4](#) I present a standard algorithm for causal discovery: the PC algorithm. I then illustrate by means of a simple abstract example how this algorithm works when applied to variable sets containing variables standing in causal as well as in constitutive relevance relationships. Whenever the algorithm outputs an edge between two variables, the question arises whether this edge stands for a causal or a constitutive relationship. I suggest to use information about time order to answer such questions. In [section 5](#) I show how time order information together with part-whole relationship knowledge can be used to test the empirical adequacy of my suggestion to formally treat constitution like causation. I also discuss several limitations of the proposed search procedure and try to resolve possible worries philosophers might have concerning violations of the causal Markov condition and the causal faithfulness condition in models of mechanisms. I conclude in [section 6](#).

2 Constitutive relevance relations in mechanisms

Mechanisms are systems whose behaviors can be described at two different levels: the macro level (i.e., the mechanism as a whole) and the micro level (i.e., the mechanism's parts). How some (but typically not all) of the parts at the micro level behave and causally interact with each other is responsible for the macro behavior of the system as a whole. The behaviors of these micro parts constitute (or are constitutively relevant for) the behavior of the system as a whole. In order to give a mechanistic explanation of some macro phenomenon, one has to distinguish its constitutively relevant parts from the ones which are not constitutively relevant and provide detailed information about the causal

relations among these constitutively relevant parts. This paper is about how the constitutively relevant micro parts of a mechanism can be identified.

Most of the time I will represent the behaviors of whole systems and their parts by means of variables and their possible values. So X_1, \dots, X_n could, for example, describe the possible behaviors of the parts of a car, and Y could be a macro variable with two possible values: y_1 for the car is moving and y_0 for the car is not moving. Now we can ask two questions. First, we can ask which instantiations x_1, \dots, x_n of the variables X_1, \dots, X_n constitute the specific behavior y_1 (i.e., the car is moving) and which constitute behavior y_0 (i.e., the car is not moving). But we can also be more abstract and shift the question from the value to the variable level. When doing this, we ask which of the variables X_1, \dots, X_n are constitutively relevant for Y . There is, of course, a close connection between these two questions: Whenever X_1, \dots, X_n are constitutively relevant for Y , at least some values x_i of every micro variable X_i have to be constitutively relevant for some behavior y of the system as a whole, and whenever some values x_i of one of the micro variables X_i are constitutively relevant for some Y -value y , then X_i will also be constitutively relevant for Y . In this paper, I am mainly interested in constitutive relevance between variables X_1, \dots, X_n and Y .

How the behaviors modeled depend on each other will be represented by a probability distribution P over the set of variables of interest.¹ Let us now ask how P has to be constrained in case it is a distribution over a set of variables

¹Note that there are several possible ways that probability distributions over sets of variables representing mechanisms can obtain. There might, for example, be systems whose constituent variables do not change their values over time. In that case one gets probability distributions by looking at the different values variables take in spatiotemporally different systems of similar type. An example would be probability distributions over mineral components of rocks (cf. Ramsey et al., 2002). The other possibility is that constituent variables change their values over time in one system (see, e.g., Chu, Glymour, Scheines, & Spirtes, 2003). I would like to thank an anonymous referee for pointing this out to me.

$\{X_1, \dots, X_n, Y\}$ in which X_1, \dots, X_n are all the constituents of Y in a mechanism. To answer this question, we have to take a closer look at the characteristic marks of mechanisms. There are several controversies about these characteristic marks in the literature on mechanisms. It is, for example, controversial whether the macro behavior of a mechanism can be ontologically reduced to the micro behaviors of its constitutively relevant parts (see, e.g., [Eronen, 2011](#); [Kistler, 2009](#); [Fazekas & Kertesz, 2011](#); [Soom, 2012](#)). There is also no consensus about whether constitutive relevance is a relation totally different from causation or a special form of causal dependence. Most mechanists, headed up by [Craver and Bechtel \(2007\)](#), support the view that constitution is a non-causal notion. [Leuridan \(2012\)](#), however, convincingly showed that if one characterizes constitutive relevance by Craver's ([2007b, 2007a](#)) mutual manipulability account, which explicitly endorses Woodwardian ([2003](#)) ideal interventions, then it follows that constitutive relevance is a special form of causation. (From this observation, of course, one could also infer that mutual manipulability is not suited for accounting for constitutive relevance or that Woodward's theory of causation has limitations when applied to mechanisms.)

In this paper I prefer to stay neutral on controversial questions like the ones described above. The method for uncovering constitutive relevance relations in mechanisms that I will propose is compatible with macro phenomena being ontologically reducible or not ontologically reducible to their constitutively relevant micro parts. It is also compatible with constitutive relevance being a causal relation or a non-causal relation. I will now focus on the uncontroversial characteristic marks of mechanisms and constitutive relevance relations. Note that I do not intend to explicitly define constitutive relevance. I rather try to describe some of its uncontroversial characteristics, or in Glymour's ([2004](#)) words: I endorse a Euclidean instead of a Socratic account of constitutive relevance.

It is uncontroversial that a mechanism's macro behavior supervenes on the behaviors of its constitutively relevant micro parts (Craver, 2007b, p. 153), meaning that every change in the macro behavior is necessarily accompanied by a change in the behavior of some constitutively relevant micro parts. For our representation by means of variables and probability distributions, this implies that conditionalizing on different values of the macro variable Y has to lead to a probability change for some instantiation x_1, \dots, x_n of the variables X_1, \dots, X_n constitutively relevant to Y . (For the moment let us assume that X_1, \dots, X_n are all the variables constitutively relevant to Y .) Hence, the following property obtains due to supervenience:

Supervenience $\forall y, y' \exists x_1, \dots, x_n : y \neq y' \rightarrow P(x_1, \dots, x_n | y) \neq P(x_1, \dots, x_n | y')$

While the behavior of a mechanism at the macro level supervenes on the behaviors of its constitutively relevant parts, every combination of behaviors of these parts constitutes a certain macro state of the mechanism. Once we know in detail how the constitutively relevant parts behave, we can predict the behavior of the system as a whole with certainty. For our representation by means of variables and probability distributions this means that there is also the following constraint on the probability distribution P over a set of variables $\{X_1, \dots, X_n, Y\}$ fully describing a mechanism's macro behavior and the behaviors of its constitutively relevant parts. When we conditionalize on any instantiation x_1, \dots, x_n of the micro variables X_1, \dots, X_n , then the probability of a certain Y -value y is 1:

Constitution $\forall x_1, \dots, x_n \exists y : P(y | x_1, \dots, x_n) = 1$

The last typical feature of mechanisms I want to discuss here is multiple realizability. In a mechanism, the macro behavior of the system as a whole is

typically (but not necessarily) multiply realizable by the behaviors of its constitutively relevant parts. This means that while every combination of instantiations of the constitutively relevant micro variables fully determines the system's macro behavior, the diverse macro behaviors of a mechanism are compatible with more than just one combination of instantiations of the micro variables. In terms of probabilities, this amounts to the statement that the conditional probabilities $P(x_1, \dots, x_n | y)$ have to be smaller than 1:

Multiple realizability $\forall x_1, \dots, x_n \forall y : P(x_1, \dots, x_n | y) < 1$

Let me briefly introduce the following abstract toy example (see [Figure 1](#)) for further illustration. Our system of interest can be described by the ten variables X_1, \dots, X_{10} . The variables X_1, \dots, X_8 describe behaviors at the micro level; the variables X_9 and X_{10} describe behaviors at the macro level. The objects whose behaviors are modeled by X_1, \dots, X_4 are parts of the objects whose behaviors are represented by X_9 . The objects whose behaviors are modeled by X_6 and X_7 are parts of the objects whose behaviors are represented by X_{10} . These part-whole relationships are indicated by the ellipses in [Figure 1](#). The direct causal influences of the micro variables on each other are indicated by the continuous arrows in [Figure 1](#), where the variables at the arrows' tails are direct causes of the variables at the arrows' heads w.r.t. the set of all variables. Let us assume that we are only interested in causal relations among micro variables. Hence, we draw no continuous arrows exiting or pointing at one of the macro variables X_9 and X_{10} . This also nicely fits the idea of mechanistic explanation, which requires only information about a macro phenomenon's constitutively relevant parts and their causal connections.

Now let us also assume that X_1 , X_2 , and X_3 (but not X_4) represent the behaviors of the parts constitutively relevant for X_9 , and that X_6 and X_7 are

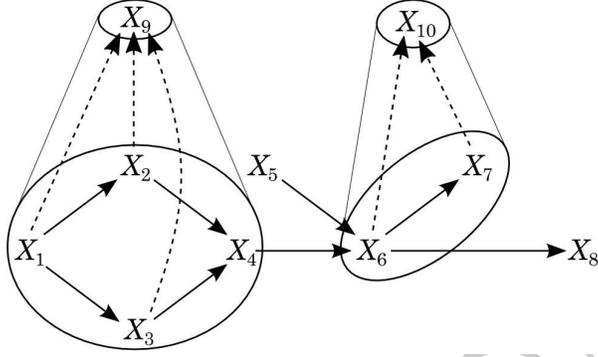


Figure 1

constitutively relevant for X_{10} . Constitutive relevance relationships are indicated by the dashed arrows in Figure 1. In addition, let us assume that X_1 , X_2 , and X_3 are all the constitutively relevant parts for X_9 and that X_6 and X_7 are all the constitutively relevant parts for X_{10} . In that case, X_9 supervenes on $\{X_1, X_2, X_3\}$ and X_{10} supervenes on $\{X_6, X_7\}$, meaning that for all X_9 -values x_9 and x'_9 with $x_9 \neq x'_9$ there are at least some value instantiations x_1, x_2, x_3 of the micro variables X_1, X_2, X_3 such that $P(x_1, x_2, x_3|x_9) \neq P(x_1, x_2, x_3|x'_9)$ holds, and that for all X_{10} -values x_{10} and x'_{10} with $x_{10} \neq x'_{10}$ there are at least some value instantiations x_6, x_7 of the micro variables X_6, X_7 such that $P(x_6, x_7|x_{10}) \neq P(x_6, x_7|x'_{10})$ holds. Furthermore, X_1, X_2, X_3 constitute X_9 and X_6, X_7 constitute X_{10} . This means that for every instantiation x_1, x_2, x_3 of the micro variables X_1, X_2, X_3 there is a certain X_9 -value x_9 such that $P(x_9|x_1, x_2, x_3) = 1$ holds, and that for every instantiation x_6, x_7 of the micro variables X_6, X_7 there is a certain X_{10} -value x_{10} such that $P(x_{10}|x_6, x_7) = 1$ holds. Finally, we also assume multiple realizability, i.e., that no X_9 -value determines any instantiation of X_1, X_2, X_3 with probability 1 and that no X_{10} -value determines any instantiation of X_6, X_7 with probability 1.

3 Constitutive relevance and causal Bayes nets

A causal Bayes net is a triple $\langle V, E, P \rangle$. V is a set of variables, $G = \langle V, E \rangle$ is a directed acyclic² graph, and P is a probability distribution over V . The arrows in G indicate direct causal connections w.r.t. V . The variables $X \in V$ with $X \rightarrow Y$ in G are called Y 's parents. The set of Y 's parents is referred to as $Par(Y)$. The variables X that are connected to Y via a chain of arrows of the form $X \rightarrow \dots \rightarrow Y$ are called Y 's ancestors. The set of Y 's ancestors in a graph is referred to as $Anc(Y)$. The variables Y that are connected to X via a chain of arrows of the form $X \rightarrow \dots \rightarrow Y$, on the other hand, are called X 's descendants. For technical reasons every variable X is assumed to be a descendant of itself. $Des(X)$ is the set of all descendants of X . Finally, The probability distribution P represents the strengths of the influences propagated along the causal arrows.

CBNs are assumed to satisfy the causal Markov condition (Spirtes et al., 2000, p. 29):

Definition 3.1 (causal Markov condition). $\langle V, E, P \rangle$ satisfies the causal Markov condition if and only if every variable $X \in V$ is probabilistically independent of its non-descendants conditional on its parents.

CMC is basically a generalisation of two insights which can be traced back to Reichenbach's (1956) influential book *The direction of time*: Conditionalizing on all common causes of two correlated variables screens these variables off each other (given there are no other causal connections between these two variables), and conditionalizing on all direct causes of a variable screens this variable off from its indirect causes. Informally, the causal Markov condition says that all correlations among variables in a variable set V can be accounted for by causal

²A graph is acyclic if it does not feature a path of the form $X \rightarrow \dots \rightarrow X$.

connections among variables in V (cf. [Schurz & Gebharder, 2016](#)).

If $\langle V, E, P \rangle$ satisfies the causal Markov condition, then the graph $G = \langle V, E \rangle$ determines the following Markov factorization ([Spirtes et al., 2000](#), pp. 29f):

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Par}(X_i)) \quad (1)$$

CBNs can also satisfy several additional conditions. For this paper, only CMC and the causal faithfulness condition (CFC) will be relevant. While CMC only implies independencies for a given causal graph, CFC implies dependencies given a causal structure. One formulation of CFC goes as follows ([Zhang & Spirtes, 2008](#), p. 247):

Definition 3.2 (causal faithfulness condition). *$\langle V, E, P \rangle$ satisfies the causal faithfulness condition if and only if the independencies implied by CMC and $G = \langle V, E \rangle$ are all the independencies in P .*

In the argumentation for why constitutive relevance implies the same screening off relations as direct causal connection, I will assume the causal Markov condition to hold. In the argumentation for why constitutive relevance often comes with the same linking-up property as direct causal relations, I assume the causal Markov and the causal faithfulness condition to be satisfied. I will discuss possible failures of Markov and faithfulness in models of mechanisms in [section 5](#).

Let us now ask how constitutive relevance relations can be represented within a CBN. First of all, note that in the causal modelling literature it is typically assumed that the sets of variables we are interested in do not contain variables standing in non-causal relations. Such non-causal dependencies are, for example, dependencies due to definition, conceptual dependence, logical dependence, dependence due to part-whole relationships, and also dependencies of superven-

nience and constitution. The reason for excluding such dependencies is that this move guarantees that all correlations among variables in a variable set V of interest have to be due to causal relations. So we can directly infer a causal dependence of some sort when observing a probabilistic dependence.

For this paper, however, we also allow our variable sets to contain variables standing in constitutive relevance relations. As [Woodward \(2015, sec. 1\)](#) remarks, the question of how to represent non-causal dependencies in a causal model is not trivial. In the remainder of this section I will argue that constitutive relevance relations can, in principle, be formally treated as a special kind of causal relations in CBNs. Hence, we can apply the formal apparatus introduced above when analyzing variable sets with variables which may stand in causal as well as in constitutive relationships. Note that I do not want to claim that constitutive relevance is a form of causation. I prefer to stay neutral on that question. My claim is rather that the formal tool of CBNs can be used to capture constitutive relevance relations the same way it can be used to capture a certain kind of causal dependence.

To see why constitution can be represented similarly to direct causal connection in a CBN, we have to take a closer look at how arrows in CBNs work. Causal arrows in a CBN possess a certain robustness or stability property. Since joint probabilities factor according to [Equation 1](#), the conditional probabilities $P(X_i|Par(X_i))$ —these conditional dependencies correspond to the arrows from X_i 's parents to X_i and are also called X_i 's parameters—cannot be changed by varying the prior distribution of non-descendants of X_i . This robustness of the causal arrows in a CBN nicely fits the idea that the variables X_i of a CBN together with their parents $Par(X_i)$ stand for autonomous causal mechanisms (cf. [Pearl, 2000](#), p. 22).

Now notice that constitutive relevance relations in mechanisms also seem to

possess this stability property. Assume X_1, \dots, X_n are all the variables constitutively relevant for Y . Then every instantiation x_1, \dots, x_n of X_1, \dots, X_n forces Y to take a particular value y (with probability 1). All other Y -values y' get probability 0 when conditionalizing on x_1, \dots, x_n . Since $Y = y$ is constituted by x_1, \dots, x_n , the conditional probabilities $P(y|x_1, \dots, x_n)$ will not change when changing the prior distribution over some non-descendants of Y , meaning that constitutive relevance relations share the same stability property as direct causal relations. In the case that X_1, \dots, X_n are all the variables constitutively relevant to Y , then constitution behaves like deterministic causation. If, on the other hand, X_1, \dots, X_n are not all the variables constitutively relevant to Y , then it is not guaranteed that every instantiation x_1, \dots, x_n of X_1, \dots, X_n forces Y to take a particular value y with probability 1. However, X_1, \dots, X_n will still provide the maximal probabilistic information among non-descendants of Y available in our variable set V . And that is all required for constitutive relevance relations to be representable by arrows in a CBN.

There are connected and additional reasons for treating constitutive relevance relations like direct causal connections in a CBN. One is that constitutive relevance relations seem to share the same screening off and linking-up properties as causal relations: If the only causal connection between two variables X and Y is a directed path through Z ($X \rightarrow Z \rightarrow Y$ or $X \leftarrow Z \leftarrow Y$) or a common cause path through Z ($X \leftarrow Z \rightarrow Y$), then X and Y will be independent when conditionalizing on any Z -value. This is a direct consequence of CMC. But if X and Y are, on the other hand, only connected through a common effect Z ($X \rightarrow Z \leftarrow Y$), then X and Y will become dependent when conditionalizing on certain Z -values. This is a direct consequence of CFC.

Let me briefly illustrate that constitutive relevance relations also share these screening off and linking-up properties. Let us start with screening off: Assume

Z is constituted by Y_1, \dots, Y_m , and each Y_i (with $1 \leq i \leq m$) is constituted by some variables X_1, \dots, X_n . Let us further assume that Y_1, \dots, Y_m are all the variables constitutively relevant to Z and that the variables X_1, \dots, X_n are all the variables constitutively relevant to the variables Y_i . In that case constitutive relevance will, again, behave exactly like deterministic causation: Conditionalizing on a value combination y_1, \dots, y_m of instantiations of Y_1, \dots, Y_m will screen Z off from any of the mentioned variables X_1, \dots, X_n . For Z 's value it is not important which values of X_1, \dots, X_n realize the values of Y_1, \dots, Y_m we are conditionalizing on. If we have missed some of the constitutively relevant variables we would still get the same screening off effect, simply because, again, the maximally available probabilistic information for every variable among its non-descendants is already provided by its constituents captured by our variable set V . So we observe the same screening off properties in constitutive chains as in causal chains. Similar considerations apply to the case where Z is constitutively relevant for both X and Y . First, X and Y are dependent because they have a common constitutively relevant part, viz. Z . If one now conditionalizes on Z , then this correlation will break down. Once Z 's value is fixed, X 's state will not give us any additional information about Y 's state and vice versa.

Let us now come to the linking-up property and briefly consider the following toy example: Assume that we are interested in the behavior of a bidder in an auction. The macro behavior of interest is whether the bidder bids or does not bid. Bidding ($B = 1$) is constituted by raising the left hand ($L = 1$ and $R = 1$ as well as $L = 1$ and $R = 0$) as well as by rising the right hand ($R = 1$ and $L = 1$ as well as $R = 1$ and $L = 0$). Not bidding ($B = 0$) is constituted by raising neither the left nor the right hand ($L = 0$ and $R = 0$). Note that L and R are fully independent. Whether the bidder raises one or both of her hands depends only on the bidder's own free decision. Now we can observe that conditionalizing on

$B = 1$ will render L and R dependent: If $B = 1$, then the probability of $L = 1$ will be greater than 0, but smaller than 1; but if we learn in addition that $R = 0$, then this will increase the probability of $L = 1$ to 1. Hence, constitutive relations between several variables X_1, \dots, X_n constituting Y seem to behave exactly like they were direct causal relations from X_1, \dots, X_n to Y .

Another reason for treating constitutive relevance relations in mechanisms like direct causal connections in CBNs is that this nicely fits the idea of multiple realizability: While the probability distribution over a macro variable Y constituted by X_1, \dots, X_n cannot change when X_1, \dots, X_n are fixed to some values x_1, \dots, x_n , fixing Y to one of its values does—at least in principle—not determine the values of X_1, \dots, X_n , meaning that different value combinations of X_1, \dots, X_n are—at least in principle—compatible with Y 's taking some value y .

4 Uncovering constitutive relevance relations

Standard search procedures for causal dependence (as, for example, presented in [Spirtes et al., 2000](#)) output a structure consisting of different kinds of edges between variables. These different kinds of edges encode different kinds of causal information. If my argumentation in [section 3](#) is correct, then constitutive relevance relations behave formally like direct causal relations in CBNs in so far as they produce the Markov factorization. Hence, applying standard procedures for causal discovery to variable sets containing variables which do not only stand in causal, but also in constitutive relationships to each other, should lead to a structure in which each edge either indicates a causal or a constitutive dependence. Let me illustrate this by applying what is probably the most well-known algorithm for causal discovery to the abstract example introduced earlier: the PC algorithm ([Spirtes et al., 2000](#), pp. 84f). In the following, $Adj(G, X)$ is the set of variables adjacent to X in graph G , i.e., the set of variables Y

with $X \longrightarrow Y$, $Y \longrightarrow X$, or $X - Y$ in G . Note that the graph G as well as the sets $Adj(G, X) \setminus \{Y\}$ and $Sep(X, Y)$ are constantly updated when going through steps S1 to S4 of the algorithm.

PC algorithm

S1: Form the complete undirected graph G on the vertex set V .

S2: $n = 0$.

repeat

repeat

select an ordered pair of variables X and Y that are adjacent in G such that $Adj(G, X) \setminus \{Y\}$ has cardinality greater than or equal to n , and a subset S of $Adj(G, X) \setminus \{Y\}$ of cardinality n , and if X and Y are independent conditional on S delete edge $X - Y$ from G and record S in $Sep(X, Y)$ and $Sep(Y, X)$;

until all ordered pairs of adjacent variables X and Y such that $Adj(G, X) \setminus \{Y\}$ has cardinality greater than or equal to n and all subsets S of $Adj(G, X) \setminus \{Y\}$ of cardinality n have been tested for independence;

$n = n + 1$;

until for each ordered pair of adjacent vertices X and Y , $Adj(G, X) \setminus \{Y\}$ is of less than n .

S3: For each triple of vertices X, Y, Z such that the pair X, Y and the pair Y, Z are each adjacent in G but the pair X, Z are not adjacent in G , orient $X - Y - Z$ iff Y is not in $Sep(X, Z)$.

S4: repeat

if $X \rightarrow Y$, Y and Z are adjacent, X and Z are not adjacent, and there is no arrowhead at Y , then orient $Y - Z$ as $Y \rightarrow Z$;

if there is a directed path from X to Y , and an edge between X and Y , then orient $X - Y$ as $X \rightarrow Y$;

until no more edges can be oriented.

This algorithm presupposes CMC, CFC, and acyclicity. It produces a structure which may feature directed as well as undirected edges. If our set of variables V would contain only variables not standing in non-causal dependencies (such as constitutive relevance relations), then a directed edge from a variable X to another variable Y ($X \rightarrow Y$) would stand for X 's being a direct cause of Y w.r.t. V , while an undirected edge between X and Y ($X - Y$) would stand for X 's being a direct cause of Y or Y 's being a direct cause of X w.r.t. V . If we also allow for constitutive relevance relations among variables in V , on the other hand, arrows and edges can stand for direct causal dependence or for direct constitutive dependence. Figure 2 illustrates—by means of our exemplary system introduced earlier—how the PC algorithm works in case our exemplary system is Markov and faithful. In that case, PC outputs the structure depicted in Figure 2(d).

As a final step, we have to distinguish edges standing for direct causal dependencies from edges representing constitutive relevance relations. From earlier considerations we already know that constituted variables Y must be descendants of their constitutively relevant variables X_1, \dots, X_n and that the presence of all variables X_1, \dots, X_n constitutively relevant to a macro variable Y in our variable set V forces certain constraints on the system's probability distribution P (given all the constitutively relevant variables are in V , which we

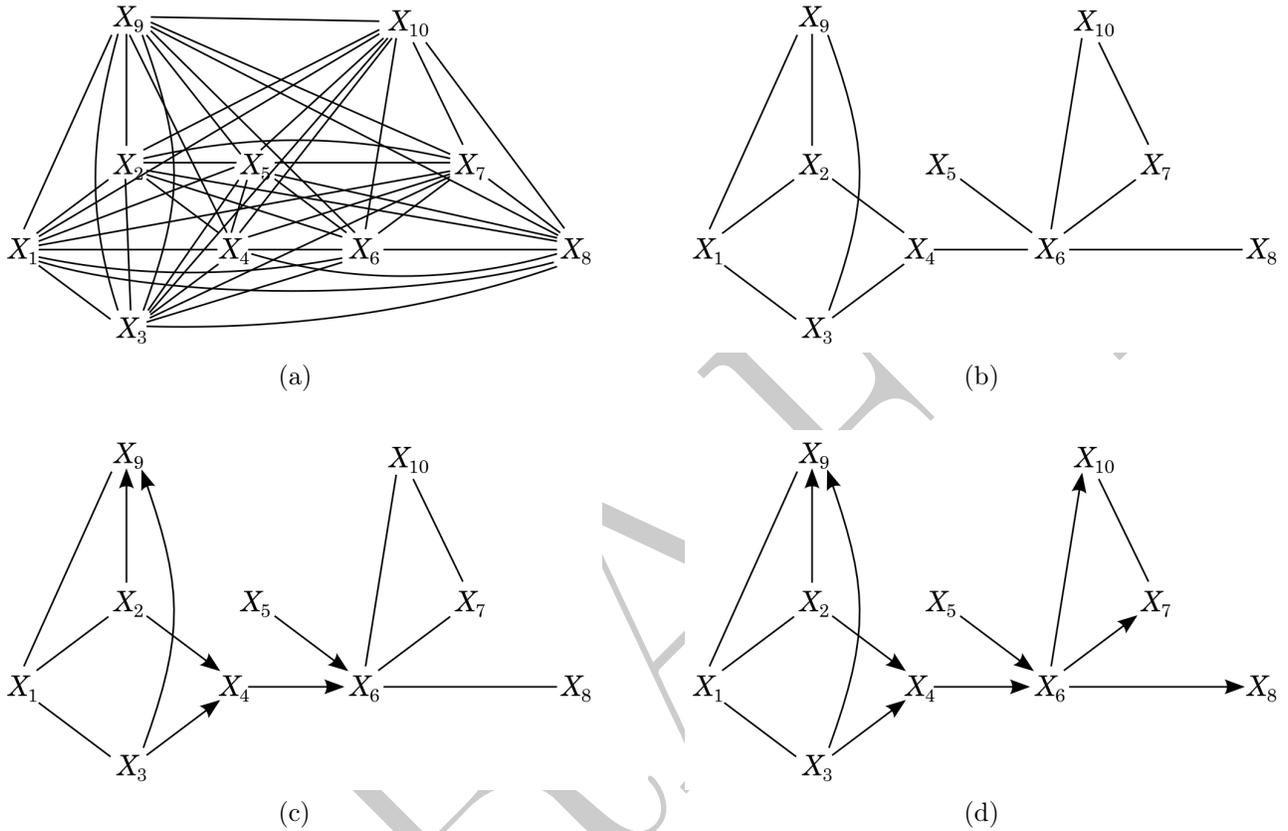


Figure 2: (a) shows the graph one gets after applying step 1. Here we simply connect every pair of variables by an undirected edge. In step 2 we sort out all those edges which do not represent direct causal (or constitutive) relationships among variables in V . The graph resulting from step 2 is shown in (b). In step 3 we systematically detect common effects (or constituted macro variables) in our system. (c) shows the structure resulting from step 3. Finally, step 4 allows us to orient some additional undirected edges. (d) shows the graph the PC algorithm outputs after step 4.

assumed in our simple exemplary system). In particular, the conditional probabilities $P(x_1, \dots, x_n|y)$ for some instantiations x_1, \dots, x_n of the micro variables X_1, \dots, X_n have to differ from $P(x_1, \dots, x_n|y')$ for all $y' \neq y$ due to the fact that macro variables supervene on their constitutively relevant micro variables, and every combination x_1, \dots, x_n of values of variables X_1, \dots, X_n has to determine Y to take a certain value y due to the fact that X_1, \dots, X_n constitute Y . However, these constraints could—at least in principle—also be satisfied by a purely causal common effect structure. So we should not read these constraints as guarantee for X_1, \dots, X_n being constitutively relevant for Y .

An alternative way to go in order to distinguish causal arrows from constitutive arrows in [Figure 2\(d\)](#) is to use information about time order. In mechanisms, causes always precede their effects in time, while constitutively related events can be expected to regularly occur simultaneously. More precisely, we assume that the time intervals at which the instantiations of a variable X occur are always strictly before the time intervals at which the instantiations of a variable Y occur if X is a cause of Y . If X and Y are, on the other hand, connected by a constitutive relevance relation, then the time intervals at which X -instantiations occur should regularly overlap with the time intervals at which Y -instantiations occur. Note that this is a very weak constraint that probably almost all mechanists want to endorse.

With this constraint in mind, we can now simply use time order information to decide which of the edges the PC algorithm outputs represent constitutive relevance relations. In the exemplary system introduced in [section 2](#), the time intervals at which instantiations of X_9 occur will regularly overlap with the time intervals at which X_1 -, X_2 -, and X_3 -instantiations occur. From this we can infer that the undirected edge $X_1 - X_9$ and the arrows $X_2 \rightarrow X_9$ and $X_3 \rightarrow X_9$ cannot indicate direct causal dependence, but have to stand for constitutive rel-

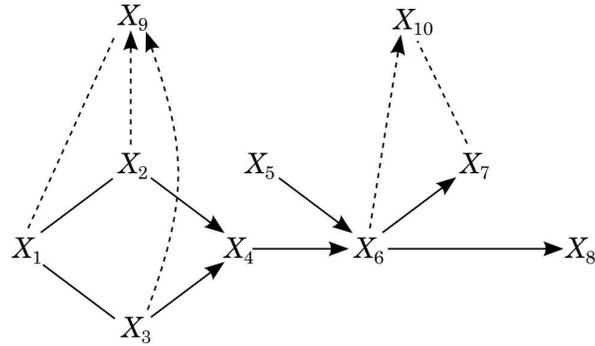


Figure 3

evance relationships. A similar conclusion we can draw for the arrow $X_6 \rightarrow X_{10}$ and the undirected edge $X_7 - X_{10}$. To graphically distinguish edges representing constitutive relations from causal relations, let us draw the former as dashed and the latter as continuous edges. So we arrive at the structure in Figure 3. We were finally able to correctly identify all the constitutive relevance relations and also all the causal relations among micro variables of our system of interest. In addition, we were able to orient most of the structure resulting from PC's edges, which is also nice.

At first glance it might seem as if PC plus time order information cannot output continuous arrows from micro variables to macro variables. There are, however, possibilities that micro variables can be output as causes of macro variables.³ This can be the case when there is a clear time lag between the events described by a micro variable X and a macro variable Z . This might happen, for example, if the set of variables V does contain a micro cause X of a constituent variable Y of a macro variable Z , but not the constituent variable Y itself. In that case PC might output a continuous edge between X and the variable Z .

³I would like to thank an anonymous referee for pushing me to think about this issue.

5 Applicability, limitations, and empirical adequacy

In the case of our exemplary causal and constitutional system and under the assumption that it is Markov and faithful, the discovery procedure suggested works well for uncovering the constitutive relevance relations as well as the causal relations among micro variables featured by this system. The system was, of course, designed to show how powerful constitutive relevance discovery on the basis of PC can be. Not all systems can be expected to be that nice. Now the question that naturally arises is the following: Are there many systems involving mechanisms that can be expected to meet the requirements for a successful application of PC? Many philosophers would probably immediately object that it is well-known that systems featuring variables standing in other than causal dependencies are usually not Markov. In addition, models of mechanisms can be expected to feature deterministic dependencies due to the fact that a mechanism's components determine its macro level. But it is also well-known that systems featuring deterministic dependencies will most of the time violate the faithfulness condition. So at first glance it may seem that there might be almost no systems (involving mechanisms) that satisfy the requirements for running PC. I discuss these concerns and connected questions about the applicability and limitations of the suggested discovery procedure in the next paragraphs. Finally, I will discuss the empirical adequacy of my claim that constitutive relevance relations can be represented in CBNs the same way as direct causal connection can be tested.

Worries concerning the Markov condition There are basically three worries one might have with regards to the causal Markov condition in connection to mechanisms. The first one of these worries can be illustrated by means of the

following simple example: Assume we are interested in a variable set $V = \{X, Y\}$. Let us further assume that X is constitutively relevant to Y and that X and Y are probabilistically dependent. In that case, CMC demands one of the causal arrows $X \rightarrow Y$ or $Y \rightarrow X$ to be part of our model. However, since neither X is a cause of Y nor Y is a cause of X , we are not supposed to draw such a causal arrow in our model and CMC is violated. Generalized: The problem seems to be that if we do not represent non-causal dependencies with arrows, then we might end up with correlations that are not accounted for by some paths in our models' graphs.

If my earlier argumentation is correct, however, we can represent constitutive relevance with the arrows of a Bayes net in the same way as direct causal relations. We can treat this particular non-causal dependence as it were some kind of causal dependence. This means that in the problematic toy example introduced in the paragraph above, we can simply draw an arrow from the micro variable X to the macro variable Y and the resulting model will be Markov again. One might object that the *causal* Markov condition requires all arrows to be interpreted as direct *causal* relations and that we are not allowed to draw an arrow that represents a non-causal relation. My response would be that this is not so clear. If constitutive relevance behaves like a special form of causation, then why should we not represent this relation the same way as we represent an ordinary causal relation, viz. by an arrow? From a formal point of view this seems perfectly legitimate. Alternatively, one could formulate a new causal or constitutive Markov condition for systems featuring causal as well as constitutive relevance relations. I prefer to stick with the first strategy here.

Here comes the second possible worry with the Markov condition: We typically assume that the Markov condition is satisfied for systems whose variable sets V contain all common causes of variables in V . Variable sets V that satisfy

this requirement are called *causally sufficient* (Spirtes et al., 2000, p. 22). If a variable set V is not causally sufficient, then one can expect violations of the Markov condition due to correlations that are not accounted for by causal connections among variables in V .⁴ Now it seems that in the case of mechanisms our variable sets V of interest will almost never be causally sufficient. The reason is that a mechanism’s macro behavior supervenes on the behaviors of its constitutively relevant parts. Hence, every cause X of a variable representing a mechanism’s macro behavior will, at the same time as it influences this macro variable, lead to a change of the probability distribution over some of the variables contained in that macro variable’s supervenience base. To account for this probability change, X has to be a common cause of the macro variable and at least one of its constitutively relevant variables.⁵ This seems to imply the following: Whenever at least one cause of a mechanism’s macro variable is not captured by our model’s set of variables V , this cause will be a common cause of that macro variable and some of its constitutively relevant variables in V and, hence, V can be expected to not be causally sufficient. Without causal sufficiency, however, the Markov condition can be expected to be violated. Hence, it seems that almost all models of mechanisms will violate the Markov condition.

What can we reply to this latter objection? First of all, causal sufficiency is intended to exclude one particular failure of causal Markov: If V is causally sufficient, then there will be no dependencies among variables in V (produced

⁴Here is a simple example: Let V be $\{X, Y, Z\}$. Our system’s causal structure is $X \rightarrow Y \rightarrow Z$. CMC demands that X and Z are screened off by the intermediate cause Y . But now assume that Y and Z have a common cause C that is not captured by V . In that case conditionalizing on Y will activate the path $X \rightarrow Y \leftarrow C \rightarrow Z$ and X and Z might still be dependent when conditionalizing on Y . In that case, our model cannot account for the dependence of X on Z given Y and CMC would be violated.

⁵Baumgartner and Gebharder (2016) have shown that a similar consequence arises within an interventionist framework such as Woodward’s (2003).

by latent common causes) that cannot be accounted for by the causal (or, in our case, causal and constitutional) connections among variables in V . Adding all latent common causes C_i of variables in V to V is, however, just one strategy to avoid failures of causal Markov due to latent common causes. Another strategy is to fix all latent common causes' values (cf. [Spirtes et al., 2000](#), p. 22). If the values of all latent common causes C_i are fixed, then causal paths $X \leftarrow C_i \rightarrow Y$ (with $X, Y \in V$) become blocked and will not produce dependencies between X and Y anymore. Does this solve our problem? Unfortunately, the answer to this question is negative. When measuring the probability distribution over a set of variables representing a mechanism one might often not be able to fix all latent common causes of the mechanism's macro variable and some of its micro variables. So it still seems that most models of mechanisms we build will violate the causal Markov condition.

Luckily, there is also a third way that failures of causal Markov due to latent common causes can be avoided. All that is required to avoid such failures is that paths $X \leftarrow C_i \rightarrow Y$ (with $X, Y \in V$ and $C_i \notin V$) do not produce a probabilistic dependence between X and Y . A path $X \leftarrow C_i \rightarrow Y$ will not lead to a dependence between X and Y if at least one of the arrows of this path is not productive, i.e., does not propagate probability in any circumstances. Productivity of single causal arrows can be defined as follows (cf. [Schurz & Gebharter, 2016](#), p. 1087):

Definition 5.1 (productivity). *An arrow $X \rightarrow Y$ in a causal model $\langle V, E, P \rangle$ is productive if and only if there is an instantiation r of $\text{Par}(Y) \setminus \{X\}$ such that X and Y are dependent conditional on r .*

Now the crucial question is how likely it is that one of the arrows in a latent common cause path $X \leftarrow C_i \rightarrow Y$ will turn out to be unproductive when dealing with mechanisms. It comes with the representation method of

constitutive relevance relations suggested in [section 3](#) that there will always be unproductive causal arrows in such paths. Let me briefly explain why: Let us assume $X \leftarrow C \rightarrow Y_1$ is such a latent common cause path. (So $X, Y_1 \in V$ and $C \notin V$.) Assume further that X is a variable representing a mechanism's macro behavior and Y_1 is a variable representing the behavior of one of its constitutively relevant parts. Let Y_1, \dots, Y_n be variables describing the behaviors of all of the mechanism's constitutively relevant parts. Some of these micro variables Y_i (with $i \neq 1$) may be elements of our variable set of interest V , and some not. Now recall from [section 2](#) that a mechanism's macro behavior is fully determined by the behaviors of all of its constitutively relevant parts. Since X is constituted by Y_1, \dots, Y_n , the deterministic dependence of X on Y_1, \dots, Y_n cannot be broken by the influence of different causes of X or otherwise. This means that in God's full causal graph, X 's value is fully determined by the values the variables Y_1, \dots, Y_n take. If there is a cause C of X (with $C \neq Y_1, \dots, Y_n$) in this graph, then the arrow $C \rightarrow X$ will turn out to be unproductive. This result can be generalized to all models of mechanisms: There cannot be latent common cause structures $X \leftarrow C_i \rightarrow Y$ featuring only productive arrows if X describes a mechanism's macro behavior. This observation rebuts the initial worry: There will be no dependencies between macro variables and micro variables due to latent common causes in models of mechanisms.

Here comes the last one of the three possible worries about causal Markov in models of mechanisms. Every micro variable that is causally relevant for another micro variable and, at the same time, constitutively relevant for some macro variable technically works like a common cause. In addition, many of a mechanism's constitutively relevant variables can be expected to be causally relevant for other micro variables. If we fail to capture all of these causally relevant constituents, then we will get the same effect as from missing causal sufficiency:

The Markov condition will most probably be violated. This is a real limitation of PC when it comes to constitutive relevance discovery. I see no straightforward solution to this problem. One possibility for avoiding the problem is to try to add as many constitutively relevant variables to V as possible. Note that the search for variables that are candidates for constituents is not as hard as searching for latent common causes in ordinary causal models (not involving any kind of non-causal dependence). We know where to search: Only (mereological) parts of the system of interest can be constitutively relevant for that system's macro behavior.

There might, of course, still be dependencies among micro variables due to ordinary latent common causes in models of mechanisms. But a mechanism's micro variables do not stand in non-causal relation to each other and so this remaining problem is just the general problem of how to guarantee for causal sufficiency, which all causal modeling approaches have to face. So it is no special problem for the endeavor of this paper and we can ignore it for now. Let us turn now to worries about failures of faithfulness.

Worries concerning faithfulness Worries about faithfulness in mechanisms are most probably due to the fact that a mechanism's constituents determine the mechanism's macro behavior. So it seems that we have to expect deterministic dependencies in many models of mechanisms. The problem is that in the case of deterministic dependencies, faithfulness is known to be more often violated than not. Here is a simple example to demonstrate why: Assume our causal model has the structure $X \rightarrow Y \rightarrow Z$. X is a deterministic cause of Y and Y is a deterministic cause of Z . Faithfulness demands that Y and Z are dependent when conditionalizing on X . But since conditionalizing on X freezes Y to a certain value y (with probability 1), Y and Z will be screened off by X .

In the case of the exemplary causal system introduced in [section 2](#), for ex-

ample, conditionalizing on $\{X_1, X_2, X_3\}$ screens X_9 off from all other variables in V and conditionalizing on $\{X_6, X_7\}$ screens off X_{10} from all other variables in V . Since the macro variables are not causally or constitutionally connected to other variables in our exemplary system, faithfulness can be expected to hold. (At least faithfulness will not be violated because of deterministic dependencies due to constitutive relationships.) It seems, however, that applying PC to our exemplary system worked so well only because no more than two levels were involved. Mechanisms are often organized in hierarchies. A mechanism may consist of several submechanisms, which may themselves consist of several submechanisms, and so on. To demonstrate the problem that might arise in systems containing variables of more than two levels, let us assume that X_1 in our example introduced in [section 2](#) is constituted by X_{11} and X_{12} . If we would add X_{11} and X_{12} to our variable set V , then $\{X_{11}, X_{12}\}$ would screen X_1 off from the macro variable X_9 and faithfulness would be violated. This seems to be highly problematic at first glance. In many cases, we will not know which variables belong to which levels and how many levels are involved. If we have to expect violations of faithfulness in such cases, it seems that the search procedure suggested can almost never be successfully applied.

Here we find another limitation. The problem arises when we have more than two levels and all constitutively relevant variables for variables in V are also in V . However, it can be expected that in building models of mechanisms we will often not capture all the parts constitutively relevant for a mechanism's macro behavior. To be sure, we can use the following, purely empirical, test: If we have captured all constituents of some macro variable in V , then there will be some deterministic dependencies among variables in V . Via contraposition we know that when we do not find deterministic dependencies in our measured probability distribution, then there is no violation of faithfulness due to con-

stitutive dependencies over more than two levels. And even in the worst case, i.e., when we have found such a deterministic dependence, we can just remove variables from V until the deterministic dependence vanishes.

There might, finally, be violations of faithfulness in models of mechanisms that have nothing to do with the mixed causal and constitutive nature of these models. But these problems are not specific problems for models of mechanisms. The fact that such faithfulness violations can arise in models of mechanisms should not bother us more than the fact that such faithfulness violations can arise for ordinary causal structures (involving no other than causal relations).

In summary, possible problems concerning the causal Markov condition might be avoided. As long as we are able to capture all (or almost all) constituents of macro variables in V , there are no (or almost no) failures of Markov to be expected in models of mechanisms that are more drastic than what we have to expect in ordinary causal models. And as long as we know that no more than two levels are involved, the same holds with respect to faithfulness. There are no special cases of violations of Markov and faithfulness to be expected from models of mechanisms in that case. If we do not know whether more than two levels are involved, there might, however, be failures of faithfulness due to deterministic dependencies. We can avoid such failures by removing some of the constitutively relevant variables from V . Here, so I believe, lies the main limitation of the suggested search procedure. It consists in some kind of tension: In order to get Markov, we have to add as many constituents as possible to V , and in order to get faithfulness (in case of more than two levels), we are not supposed to add all variables constitutively relevant to variables in V to V .

Here are some thoughts about how one could avoid this problem when dealing with systems possibly involving more than two levels. First, we test for deterministic dependencies among variables of a set V of interest. If we find no

deterministic dependencies, then we proceed with step two. If we find deterministic dependencies, on the other hand, then we remove variables until these dependencies vanish. This should guarantee that faithfulness is not violated due to deterministic dependencies. As a second step, we run PC and use time order information to distinguish macro from micro variables. Note that our procedure will provide this information reliably (even if CMC is violated) if we were able to catch at least one constituent of every macro variable in V . Next, remove the macro variables from V . Run PC on the resulting set of micro variables. This gives us the causal structure among micro variables. We have finally learned the causal structure among micro variables as well as which micro variables are constitutively relevant for which macro variables in V .

Empirical adequacy Philosophers interested in mechanisms may find it attractive that the discovery procedure for constitutive relevance relations suggested in [section 4](#) together with additional information about part-whole relationships between the objects whose behaviors are modeled by the variables in V can be used to test the empirical adequacy of my earlier suggestion ([section 3](#)) to formally treat constitutive relevance relations similarly to direct causal relations in causal Bayes nets. If this suggestion were not empirically adequate, then we should be able to find at least some real-world systems with variables X_1, \dots, X_n and Y such that the objects whose behaviors are modeled by X_1, \dots, X_n are parts of the objects modeled by Y for which our search procedure outputs dashed arrows from Y to some of the variables X_1, \dots, X_n . More informally speaking: Our search procedure would be empirically inadequate if it even occasionally outputs dashed arrows standing for constitutive relations exiting macro variables and pointing at micro variables.

As mentioned in [section 1](#), [Ramsey et al. \(2002\)](#) use a slightly different version of PC to identify components of rock and soil samples. They found that

their procedure is able to identify the most frequently occurring kinds of carbonates equally well as or even better than human experts. If my argumentation is correct and constitutive relevance in mechanisms can be represented the same way as direct causal relations, and if constitution in mechanisms is similar in this respect to constitution in rock and soil samples—which seems to be highly probable—then their finding is strong support for the empirical adequacy of the suggested procedure for uncovering constitutive relevance relations in mechanisms.

Applications Let us now ask what the discovery procedure suggested can be used for. First of all, it allows one to do what Craver’s (2007b, 2007a) mutual manipulability approach was intended to do: Given we can describe the macro behavior of a system by a variable Y and the behaviors of all of its parts by variables X_1, \dots, X_n , then we can distinguish the constitutively relevant parts from the parts which are not constitutively relevant. In case of our exemplary system, for instance, the variables X_1, \dots, X_4 model the behavior of parts of the objects modeled by X_9 . Our search procedure correctly told us that X_1 , X_2 , and X_3 are constitutively relevant for X_9 , while X_4 is not.

In addition, the search procedure suggested has as a nice side-effect that it also allows one to distinguish macro from micro variables. If we do not already possess part-whole knowledge, i.e., we just start with a set of variables, we can learn about the part-whole relationships among variables in V . One can see this directly from our example: When running our search procedure in [section 4](#), we did not use any part-whole information. But whenever our search procedure outputs a dashed arrow from a variable X to a variable Y , then Y must be a macro variable w.r.t. X and X must be a micro variable w.r.t. Y . So it allows one to cluster variables in micro and macro variables. This, however, requires that we have already specified the set of variables we are interested in. For

an approach to infer macro variables from micro variables, see, for example, (Chalupka, Perona, & Eberhardt, 2014).

The search procedure suggested can—at least in principle—also be used when not all constitutively relevant parts are represented by variables in our vertex set V , which may sometimes be the case in practical research. To illustrate this, assume, for example, that we would have overlooked X_7 . Let V' be $V \setminus \{X_7\}$. In that case the PC algorithm would have output the same structure as in Figure 2(d), but without the subgraph $X_6 \rightarrow X_7 \rightarrow X_{10}$. So we would have been able to correctly identify all the constitutive relevance relations among variables in V' . But what if we would have overlooked X_6 instead of X_7 ? In that case CMC would probably be violated, because P_6 is a common direct ancestor of X_7 , X_8 , and X_{10} . Or, more generally: What if we do not know whether our set of variables to analyze contains all common direct ancestors of variables included in this variable set? We could use the procedure suggested directly after the discussion of possible worries concerning faithfulness in section 5. Alternatively: PC is not the only search algorithm. One could, for example, use the search procedure developed by Silva, Scheines, Glymour, and Spirtes (2006) to infer latent common ancestors and causal relations among these latent factors. Silva et al.'s algorithm, however, can only detect latent variables with at least three observed effects which are not common effects of other latent variables. One open problem is to find ways for detecting latent common causes and their causal relations requiring weaker preconditions. There are also algorithms for cyclic causal structures (see, e.g., Richardson, 1996) which are quite common in mechanisms and algorithms that only require weaker versions of CFC (see, e.g., Zhang & Spirtes, 2008). Actually, there is a multitude of search algorithms for all kinds of situations requiring different assumptions in the statistics and machine learning literature. Since the key to constitutive relevance discovery lies

in the suggestion to treat constitutional relations like causal relations, it seems safe to conjecture that many of these search algorithms supplemented by time order information can be used for uncovering constitutive relevance relations in a way similar to PC.

6 Conclusion

In this paper I investigated the question of whether constitutive relevance relations in mechanisms can—in principle—be uncovered by causal Bayes net methods. I started by providing some of the characteristic marks of constitutive relevance relations in [section 2](#). Then, in [section 3](#), I introduced the basics of the causal Bayes net formalism and argued that constitutive relevance relations in mechanisms share a certain stability property with ordinary direct causal relations: They produce the Markov factorization. I argued that, because of that, constitutive relevance can be represented within the causal Bayes net framework similarly to direct causation. In [section 4](#) I briefly introduced a standard search algorithm for causal discovery: the PC algorithm. I demonstrated by means of a simple toy example how PC would work when applied to a set of variables which do not only stand in causal, but also in constitutive relationships. The algorithm outputs a pattern in which every directed arrow or undirected edge either stands for a direct causal or a direct constitutive relationship. I then used time order information to discriminate constitutive relevance relations from causal relations: In mechanisms, the time intervals at which the micro behaviors that are constitutively relevant for a system’s macro behavior occur regularly overlap with the time intervals at which the macro behavior occurs, while causes should be expected to always occur strictly before their effects. In [section 5](#) I finally suggested a method for testing the empirical adequacy of formally treating constitutive relevance analogously to causation

in causal Bayes nets. I also discussed possible objections regarding failures of Markov and faithfulness in models of mechanisms and also some limitations of the search procedure suggested.

In summary, the search procedure should be applicable if the target system is acyclic as well as causal Markov and faithful. There are some specific limitations regarding causal Markov and faithfulness in the case of models of mechanisms. Causal Markov and faithfulness can, however, be expected to not be violated due to the special mixed causal and constitutional nature of models of mechanisms in the following two circumstances:

- The model's variables describe a system's behavior at no more than two levels and all constituents of variables in V also causally relevant for variables in V are in V .
- The model does not feature any deterministic dependencies and all constituents of variables in V also causally relevant for variables in V are in V .

I think that especially the first scenario seems promising. We know where to search for possible constituents in the case of mechanisms. Possible constituents have to be (mereological) parts of the system of interest. So it seems that there are good chances of being able to capture all (or almost all) constituents that are also causally relevant for some variables in V . We also often know from the beginning to which levels the variables of interest belong. If we are interested in more than one level and if we already have this knowledge, then we can proceed in several steps. We can start with level 1 and 2 and learn about the causal structure among micro variables at level 1 and about their constitutive relevance dependencies to variables at level 2. In a next step, we can treat the variables at level 2 as micro variables and apply the search procedure to learn about the

causal structure of the variables at level 2 and their constitutional relationships to variables at level 3 and so on.

But even if neither of the two conditions above is met or if we do not know whether one of these conditions is met, we can still apply the alternative search procedure proposed in [section 5](#): Remove variables from V until there are no deterministic dependencies among variables in V . This guarantees that there are no violations of faithfulness due to deterministic dependencies. Then run PC and use time order information to identify constitutive relevance relations and macro variables. This step will work even if CMC is violated. Remove all macro variables from V and run PC again over the resulting set of micro variables. As a result we get the causal structure among micro variables. We finally have everything required for mechanistic explanation: knowledge about the causal structure among micro variables and information about which micro variables are constitutively relevant for which macro variables.

All in all, it seems that causal Bayes net methods for causal discovery can—at least in certain circumstances—also be applied to uncover constitutive relevance relations in mechanisms. This paper can be seen as a first small step into this new research area. Which search procedures different from PC can be used in similar or different ways for constitutive relevance discovery is a project for future research.

Acknowledgements: This work was supported by Deutsche Forschungsgemeinschaft (DFG), research unit FOR 1063. My thanks go to Michael Baumgartner, Lorenzo Casini, Jens Harbecke, Beate Krickel, Gerhard Schurz, and Jon Williamson for important discussions. Thanks also to an anonymous referee for helpful comments.

References

- Baumgartner, M., & Casini, L. (in press). An abductive theory of constitution. *Philosophy of Science*.
- Baumgartner, M., & Gebharder, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *British Journal for the Philosophy of Science*, 67(3), 731–756.
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 421–441.
- Chalupka, K., Perona, P., & Eberhardt, F. (2014). Visual Causal Feature Learning. *arXiv.org*, 2309.
- Chu, T. J., Glymour, C., Scheines, R., & Spirtes, P. (2003). A statistical problem for inference to regulatory structure from associations of gene expression measurements with microarrays. *Bioinformatics*, 19(9), 1147–1152.
- Craver, C. (2007a). Constitutive explanatory relevance. *Journal of Philosophical Research*, 32, 3–20.
- Craver, C. (2007b). *Explaining the brain*. Oxford: Clarendon Press.
- Craver, C., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22(4), 547–563.
- Eronen, M. I. (2011). *Reduction in philosophy of mind*. Heusenstamm: De Gruyter.
- Fazekas, P., & Kertesz, G. (2011). Causation at different levels: Tracking the commitments of mechanistic explanations. *Biology and Philosophy*, 26(3), 365–383.
- Gebharder, A. (2015). Causal exclusion and causal Bayes nets. *Philosophy and Phenomenological Research*.
- Gebharder, A. (in press). *Causal nets, interventionism, and mechanisms: Philo-*

- sophical foundations and applications*. Dordrecht: Springer.
- Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44(1), 49–71.
- Glymour, C. (2004). Critical notice. *British Journal for the Philosophy of Science*, 55(4), 779–790.
- Glymour, C., Spirtes, P., & Scheines, R. (1991). Causal inference. *Erkenntnis*, 35(1/3), 151–189.
- Harbecke, J. (2015). The regularity theory of mechanistic constitution and a methodology for constitutive inference. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 54, 10–19.
- Illari, P. M., Russo, F., & Williamson, J. (Eds.). (2011). *Causality in the sciences*. Oxford: Oxford University Press.
- Kistler, M. (2009). Mechanisms and downward causation. *Philosophical Psychology*, 22(5), 595–609.
- Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *British Journal for the Philosophy of Science*, 63(2), 399–427.
- Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1–25.
- Pearl, J. (2000). *Causality* (1st ed.). Cambridge: Cambridge University Press.
- Ramsey, J., Gazis, P., Roush, T., Spirtes, P., & Glymour, C. (2002). Automated remote sensing with near infrared reflectance spectra: Carbonate recognition. *Data Mining and Knowledge Discovery*, 6(3), 277–293.
- Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.
- Richardson, T. (1996). A discovery algorithm for directed cyclic graphs. In

- Uai'96* (pp. 454–461). San Francisco, CA: Morgan Kaufmann.
- Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese*, 1–25.
- Schurz, G., & Gebharder, A. (2016). Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets. *Synthese*, 193(4), 1073–1103.
- Silva, R., Scheines, R., Glymour, C., & Spirtes, P. (2006). Learning the structure of linear latent variable models. *Journal of Machine Learning Research*, 191–246.
- Soom, P. (2012). Mechanisms, determination and the metaphysics of neuroscience. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(3), 655–664.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT Press.
- Woodward, J. (2003). *Making things happen*. Oxford: Oxford University Press.
- Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91(2), 303–347.
- Zhang, J., & Spirtes, P. (2008). Detection of unfaithfulness and robust causal inference. *Minds and Machines*, 18(2), 239–271.