Synthese
DOI 10.1007/s11229-010-9751-1

Intervention, determinism, and the causal minimality condition

Jiji Zhang · Peter Spirtes

Received: 12 May 2009 / Accepted: 30 October 2009

Abstract We clarify the status of the so-called causal minimality condition in the theory of causal Bayesian networks, which has received much attention in the recent literature on the epistemology of causation. In doing so, we argue that the condition is well motivated in the interventionist (or manipulability) account of causation, assuming the causal Markov condition which is essential to the semantics of causal Bayesian networks. Our argument has two parts. First, we show that the causal minimality condition, rather than an add-on methodological assumption of simplicity, necessarily follows from the substantive interventionist theses, provided that the actual probability distribution is strictly positive. Second, we demonstrate that the causal minimality condition can fail when the actual probability distribution is not positive, as is the case in the presence of deterministic relationships. But we argue that the interventionist account still entails a pragmatic justification of the causal minimality condition. Our argument in the second part exemplifies a general perspective that we think commendable: when evaluating methods for inferring causal structures and their underlying assumptions, it is relevant to consider how the inferred causal structure will be subsequently used for counterfactual reasoning.

Keywords Causation · Causal Bayesian network · Determinism · Markov condition · Intervention · Probability

J. Zhang (✉)
Lingnan University, Tuen Mun, NT, Hong Kong
e-mail: jijizhang@ln.edu.hk

P. Spirtes
Carnegie Mellon University, Pittsburgh, PA, USA

Published online: 20 June 2010

1 Introduction

It is commonplace that causation and intervention are closely related concepts, but taking the connection seriously proves to have nontrivial implications. The recent revival of the interventionist account of causation, for example, has shed new light on a variety of issues in the philosophy of causation and explanation, including especially the nature of causal explanation in special and social sciences (Woodward 2003; Steel 2006a; Campbell 2007). Besides interpretive work, the interventionist perspective also underlies a powerful framework for causal modeling and reasoning, known as causal Bayesian networks (e.g., Spirtes et al. 1993; Pearl 2000). The framework not only systemizes counterfactual reasoning based on information about a causal structure, but also stimulated a chunk of work on inference of the causal structure from data.

As we will describe in detail below, causal Bayesian networks are defined by a postulate that relates counterfactual probability distributions of a set of variables that would result from various interventions, to the actual probability distribution of the variables, mediated by a given causal structure. From the postulate one can derive interesting rules for counterfactual reasoning under various circumstances. On the other hand, a crucial component of the postulate is a condition that specifies, for a given causal structure, a set of probabilistic independence (and conditional independence) relations that must hold of a joint probability distribution if the distribution is generated from the structure. Known as the causal Markov condition, it provides a basis for causal inference based on the actual probability distribution.

Hausman and Woodward (1999, 2004a,b) presented an interventionist defense of the causal Markov condition. The details of their arguments are controversial,[1] and we do not wish to adjudicate on that dispute here. What is appealing to us, however, is their attempt to reveal an "intimate connection" between the interventionist (or manipulability) account of causation, which relates *counterfactual* circumstances, and the causal Markov condition, which relates the *actual* circumstance. To the extent that it is successful, their argument can improve our understanding of how the interventionist account of causation bears on the epistemology of causal inference in the actual circumstance. To this matter we aim to contribute two items in the present paper.

First, assuming the actual probability distribution is strictly positive, we show that the defining axiom of causal Bayesian networks—which combines the interventionist ideas of invariance and modularity with the causal Markov condition—and the interventionist interpretation of causal structure entail that the true causal structure is a *minimal* structure compatible with the actual distribution and the Markov condition. This consequence is known in the literature as the *causal minimality condition* (Spirtes et al. 1993; Hitchcock 2002), but its logical connection to the interventionist ideas seems hitherto unknown.

[1] For criticisms of their argument, see Cartwright (2002) and Steel (2006b). For a response to Cartwright, see Hausman and Woodward (2004a). Although in spirit our pursuit in the present paper can be viewed as an extension of Hausman and Woodward's work, the main point we wish to make—the point that anyone who is willing to do causal inference based on the causal Markov condition has good reason to assume the causal minimality condition as well—does not depend on the success of their arguments.

The logical connection fails when the actual probability distribution is not strictly positive, which is always the case when there are deterministic relationships between variables in the given system. Our second point is to argue, based on additional theorems without the positivity assumption, that even when the causal minimality condition is false, it would be a harmless assumption if we restricted the range of counterfactual reasoning we do with inferred causal structures. The argument exemplifies a general perspective we think commendable: when evaluating methods for inferring causal structures and their underlying assumptions, it is relevant to consider how the inferred causal structure will be subsequently used for counterfactual reasoning.

2 A proof of the causal minimality condition

Throughout this paper, we consider causal relations between variables, and to keep things simple, we assume variables under consideration are all discrete with a finite number of possible values, though the result in this section can be readily generalized to continuous variables. The causal structure for a set of variables $\mathbf{V}$ is meant to be the set of *direct causal relations* between variables in $\mathbf{V}$. Thus understood, the causal structure can be conveniently represented by a directed graph: take each variable in $\mathbf{V}$ as a vertex, and put a directed edge or arrow ($\rightarrow$) between two variables X and Y if and only if X is a direct cause of Y relative to $\mathbf{V}$. We call such causally interpreted directed graphs *causal graphs*, and we use "causal graph" and "causal structure" interchangeably in this paper.

Some graph theoretical terminology will prove useful. In a directed graph, if there is an edge $X \rightarrow Y$, X is called a *parent* of Y and Y a *child* of X. A *directed path* is an ordered sequence of two or more distinct vertices such that every vertex (except for the last one) in the sequence is a parent of its successor in the sequence. X is called an *ancestor* of Y and Y a *descendant* of X if $X = Y$ or there is a directed path from X to Y. A directed cycle occurs in the graph if there is an arrow $X \rightarrow Y$ but also a directed path from Y to X. A directed graph is called *acyclic* if there is no directed cycle in the graph. We will only consider acyclic (a.k.a. recursive) causal structures, represented by directed acyclic graphs (DAGs).

Given a DAG G over $\mathbf{V}$ and a joint probability distribution P over $\mathbf{V}$, G and P are said to be *Markov to each other* if the Markov property is satisfied: according to P, every variable is probabilistically independent of its non-descendants in G given its parents in G. The *causal* Markov condition is simply that the DAG that represents the causal structure over $\mathbf{V}$ and the joint distribution over $\mathbf{V}$ are Markov to each other.

> **Causal Markov Condition**: Given a set of variables $\mathbf{V}$ whose causal structure is represented by a DAG G, every variable in $\mathbf{V}$ is probabilistically independent of its non-effects (non-descendants in G) given its direct causes (parents in G).

This condition holds only if $\mathbf{V}$ does not leave out any variable that is a common direct cause (relative to $\mathbf{V}$) of two or more variables in $\mathbf{V}$, that is, only if $\mathbf{V}$ is *causally sufficient* (Spirtes et al. 1993). Throughout the paper we assume $\mathbf{V}$ includes enough variables to be causally sufficient.

Applying the chain rule of the probability calculus, it is easy to show that the causal Markov condition entails a factorization of the joint probability distribution P as follows:

$$P(\mathbf{V}) = \prod\nolimits_{X \in \mathbf{V}} P(X|\mathbf{Pa}_G(X))$$

where G is the DAG that represents the causal structure of $\mathbf{V}$, and $\mathbf{Pa}_G(X)$ denotes the set of X's parents in G. In words, the joint probability is a product of local pieces of conditional probability—the distribution of each variable conditional on its direct causes (for variables without a parent, it is simply the unconditional probability).

We now describe some key notions in the interventionist account of causation needed for our purpose. For any $\mathbf{S} \subset \mathbf{V}$ and a value setting $\mathbf{s}$ (a vector of values, one for each variable in $\mathbf{S}$), we assume there is a (hypothetical) intervention that sets the value of $\mathbf{S}$ to $\mathbf{s}$. We denote the intervention by $\mathbf{S} := \mathbf{s}$, and denote the counterfactual or post-intervention probability distribution that would result from the intervention by $P_{\mathrm{S}:=\mathrm{s}}$. (Stipulation: $P_{\emptyset}$ s simply the actual probability distribution P.) Clearly if the intervention $\mathbf{S} := \mathbf{s}$ does what it is supposed to do, then $P_{\mathrm{S}:=\mathrm{s}}(\mathbf{S} = \mathrm{s}) = 1$. The intervention is supposed to be *precise*, in the sense that it directly affects only its immediate targets: variables in $\mathbf{S}$. If it affects other variables at all, it is via the causal influence variables in $\mathbf{S}$ have on other variables. This precision can be achieved only if the intervention is implemented by a force *external* to $\mathbf{V}$. This external force is often modeled as an intervention variable (taking value ON or value OFF), which is supposed to be statistically independent of all those variables in $\mathbf{V}$ that are not in $\mathbf{S}$ or causally influenced by any variable in $\mathbf{S}$. (For a careful characterization of the notion of intervention, see, e.g., Woodward 2003.)

An important interventionist notion is that of *modularity* (Pearl 2000; Woodward 2003). The idea is roughly that each variable and its causal parents form a local mechanism, and each mechanism can be independently manipulated without affecting other mechanisms. In particular, if $\mathbf{S} \subset \mathbf{V}$ is manipulated, the mechanisms for variables not in $\mathbf{S}$ (i.e., variables in $\mathbf{V}\backslash\mathbf{S}$) remain unchanged.[2]

It is also very natural to interpret causal arrows in the causal structure in terms of interventions: X is a direct cause of Y relative to $\mathbf{V}$ iff there is some way of fixing all other variables in $\mathbf{V}$ to particular values, such that some change in X by intervention would be followed by a change in Y (Pearl 2000; Woodward 2003). In the present setup, the obvious formulation is:

> **Interventionist Interpretation of Direct Cause (IIDC)**: For any $X, Y \in \mathbf{V}$ and $\mathbf{Z} = \mathbf{V}\backslash\{X, Y\}$ (where '\' stands for the set-theoretical difference), X has a direct causal influence on Y relative to $\mathbf{V}$ iff there exist two values $x \neq x'$ for X and a value setting $\mathbf{z}$ for $\mathbf{Z}$, s.t. $P_{X:=x,\mathbf{Z}:=\mathbf{z}}(\mathrm{Y}) \neq P_{X:=x',\mathbf{Z}:=\mathbf{z}}(\mathrm{Y})$.

[2] Of course the notion need to be qualified—for example, some extreme intervention of a cause may well destroy the causal mechanism it figures in (Woodward 2003). Since we are interested in the consequence of the notion, and more importantly, since causal reasoning is primarily, if not exclusively, concerned with effects of normal, non-extreme interventions, we will simply assume that the interventions we consider meet the qualifications.

In plain words, X is a causal parent of Y relative to $\mathbf{V}$ if and only if, fixing all other variables to some value, there exist two interventions of X that would result in different distributions for Y.

Note that these interventionist ideas—modularity and IIDC—both involve counterfactual or post-intervention probability distributions, and so they cannot be directly appealed to in causal inference from the actual probability distribution, the distribution we can obtain samples from without having an external intervention on the causal system in question.[3] However, Hausman and Woodward (1999, 2004a,b) argued that there is an intimate connection between these ideas and the causal Markov condition,[4] which relates the causal structure to the *actual* probability distribution and hence is relevant to causal inference from the actual distribution. Is the causal Markov condition the only such inferentially relevant condition that is "intimately connected" to the interventionist ideas? We think there is at least another one.

That condition is best motivated by considering a limitation of the causal Markov condition for causal inference. Given two DAGs G and G', call G' a (proper) *subgraph* or *substructure* of G and G a (proper) *supergraph* or *superstructure* of G' if they have the same vertices, and the set of arrows in G' is a (proper) subset of the set of arrows in G. A simple well-known fact is that if a probability distribution P is Markov to a DAG G, then P is Markov to every DAG that is a supergraph of G. It follows that the causal Markov condition by itself does not warrant an inference to the absence of causal arrows.[5]

Given this limitation, the following condition should sound natural.

Causal Minimality Condition: Given a set of variables $\boldsymbol{V}$ whose actual joint distribution is P, and whose causal structure is represented by a DAG G, P is not Markov to any proper subgraph of G.

---

[3] It is usually emphasized to be *observational* or *non-experimental*. We think "actual" is a better qualifier, because there may well be some experimental control built in the causal structure under investigation. The point is that no further intervention external to the causal system of interest is involved in generating the distribution.

[4] One of their arguments (in 2004b), for example, is that conditioning on the direct causes (i.e., parents in the causal graph) of a variable simulates an external intervention on the variable (because the remaining variation after conditioning on direct causes in $\mathbf{V}$ can only be explained by external influence). So, if after conditioning on the direct causes of a variable X, the variable is still probabilistically associated with another variable Y, then it simulates the situation in which an intervention on X makes a difference to Y, and hence X should be expected to have a causal influence on Y, by the interventionist account of causation (which is a natural generalization of IIDC). In other words, if Y is not causally influenced by X, then Y should be independent of X conditional on the direct causes of X, which is the central implication of the causal Markov condition.

[5] For the simplest illustration, consider the canonical scenario of causal inference with a randomized experiment. Suppose we are interested in the causal effect of a variable X on another variable Y. We randomly assign values to X for sufficiently many samples, and observe the corresponding values of Y. The experimental setup, thanks to randomization, rules out $Y \rightarrow X$, or a common cause structure $X \leftarrow C \rightarrow Y$. If, furthermore, we establish by analyzing the data that X and Y are not probabilistically independent, we can, based on the causal Markov condition (or its well known special case, the principle of common cause), infer that X has a causal influence on Y. However, if the data analysis tells us that X and Y are probabilistically independent, the causal Markov condition alone does not license the inference to "no effect", because the structure in which there is an arrow from X to Y is still compatible with the condition.

In other words, the condition states that if the causal Markov condition is true, then the true causal structure is a minimal structure that is Markov to P. Note that the causal minimality condition is also about the *actual* probability distribution, and hence is inferentially relevant in the sense that the causal Markov condition is.

The causal minimality condition sounds like a methodological assumption of simplicity, but we shall argue that it is not just an add-on assumption, but is very well motivated in the interventionist framework. To see this, we need first motivate the defining axiom of causal Bayesian network, which specifies how a counterfactual distribution $P_{\mathbf{S}:=\mathbf{s}}$ is related to the actual distribution P according to the causal structure G. This axiom is easy to motivate given what we already said. The intervention $\mathbf{S} := \mathbf{s}$ breaks the causal arrows into the variables in $\mathbf{S}$ in the causal structure G, because variables in $\mathbf{S}$ do not causally depend on their original causal parents any more, and we have $P_{\mathbf{S}:=\mathbf{s}}(\mathbf{S} = \mathrm{s}) = 1$. For every other variable, $X \notin \mathbf{S}$, the notion of modularity imply that (1) X has the same causal parents as before the intervention; and (2) the probability distribution of X conditional on its causal parents remains the same under the intervention.

Point (1) tells us the causal structure after the hypothetical intervention $\mathbf{S} := \mathbf{s}$, which is the structure resulting from deleting all arrows into members of $\mathbf{S}$ in G. Moreover, if the causal Markov condition holds in the actual situation, we should expect it to hold in the post-intervention situation. If so, $P_{\mathbf{S}:=\mathbf{s}}$ should factorize as follows:

$$P_{\mathbf{S}:=\mathbf{s}}(\mathbf{V}) = \prod\nolimits_{S\in\mathbf{S}} P_{\mathbf{S}:=\mathbf{s}}(S) \prod\nolimits_{X\in\mathbf{V}\backslash\mathbf{S}} P_{\mathbf{S}:=\mathbf{s}}(X|\mathbf{Pa}_G(X)).$$

Moreover, given point (2) above and the fact that $P_{\mathbf{S}:=\mathbf{s}}(\mathbf{S} = \mathrm{s}) = 1$, we get the following principle[6]:

> **Intervention principle (IP)**: Given a set of variables $\mathbf{V}$ whose causal structure is represented by a DAG G, for any proper subset $\mathbf{S}$ of $\mathbf{V}$ and any value setting $\mathbf{s}$ for $\mathbf{S}$**,** the post-intervention probability distribution for $\mathbf{V}\backslash\mathbf{S}$, $P_{\mathbf{S}:=\mathbf{s}}(\mathbf{V}\backslash\mathbf{S})$, is related to the actual distribution P in the following way[7]:
>
> $$P_{\mathbf{S}:=\mathbf{s}}(\mathbf{V}\backslash\mathbf{S}) = \prod\nolimits_{X\in\mathbf{V}\backslash\mathbf{S}} P(\mathrm{X}|\mathbf{Pa}_G(\mathrm{X}))$$
>
> where $\mathbf{Pa}_G(\mathrm{X})$ is the set of parents of X in G.

The intervention principle highlights the epistemic role of causal structure: it enables calculation of counterfactual, post-intervention distributions from the actual, pre-intervention distribution. This is what causal structure is *used* for. Our argument in the next section relies heavily on this point.

[6] Spirtes et al. (1993) calls this "manipulation theorem", because it is derived from the (extended) causal Markov condition plus the invariance properties. But the derivation is almost a restatement.

[7] A more common formulation is this: the post-intervention joint distribution for $\mathbf{V}$ is related to P as follows: $P_{\mathbf{S}:=\mathbf{s}}(\mathbf{V} = \mathbf{v}) = \prod_{X\in\mathbf{V}\backslash\mathbf{S}} P(X|\mathbf{Pa}_G(X))$ when $\mathbf{v}$ is consistent with $\mathbf{S} = \mathbf{s}$; otherwise $P_{\mathbf{S}:=\mathbf{s}}(\mathbf{V} = \mathbf{v}) = 0$. This implies the formulation we use in the main text, which is more convenient for our proofs later.

One issue is that $P(X|\mathbf{Pa}_G(X))$ may be undefined (when $P(\mathbf{Pa}_G(X)) = 0$).[8] The equation in the intervention principle is understood to apply only when all the conditional probabilities are well defined. For the following theorem, we assume that the actual joint distribution P is strictly positive, i.e., every value setting for V has a non-zero probability under P, so that the conditional probabilities are all well-defined. We will see the consequence of dropping this assumption in Sect. 3.

Theorem 1 *Let* $\mathbf{V}$ *be a set of variables. Suppose its actual joint distribution* P *is strictly positive, and its causal structure as defined according to IIDC is represented by a DAG G. Then the IP implies the causal minimality condition.*

Proof Suppose the IP holds but the causal minimality condition is violated. That means there is a proper subgraph G' of G that is also Markov to P. It follows that there is an edge, say $X \to Y$, that is in G but not in G'. Since G represents the causal structure of $\mathbf{V}$ according to the IIDC, the fact that $X \to Y$ is in G implies that there exist $x \neq x'$ and $\mathbf{z}$, s.t. $P_{X:=x,\mathbf{Z}:=\mathbf{z}}(\mathrm{Y}) \neq P_{X:=x',\mathbf{Z}:=\mathbf{z}}(\mathrm{Y})$, where $\mathbf{Z} = \mathbf{V}\backslash\{X, Y\}$. However, given IP,

$$P_{X:=x,\mathbf{Z}:=\mathbf{z}}(\mathrm{Y}) = P(Y|\mathbf{Pa}_G(Y)) = P(Y|\mathbf{Pa}_{G'}(Y), X = x)$$
$$P_{X:=x',\mathbf{Z}:=\mathbf{z}}(\mathrm{Y}) = P(Y|\mathbf{Pa}_G(Y)) = P(Y|\mathbf{Pa}_{G'}(Y), X = x').$$

Since by supposition P is also Markov to G', $Y$is independent of X conditional on $\mathbf{Pa}_{G'}(Y)$ under P. This, together with the positivity of P, imply that

$$P(Y|\mathbf{Pa}_{G'}(Y), X = x) = P(Y|\mathbf{Pa}_{G'}(Y), X = x').$$

Hence $P_{X:=x,\mathbf{Z}:=\mathbf{z}}(\mathrm{Y}) = P_{X:=x',\mathbf{Z}:=\mathbf{z}}(\mathrm{Y})$, a contradiction. □

Theorem 1 suggests that, appearances to the contrary, the causal minimality condition is not just an add-on methodological preference for simplicity, but, assuming the actual probability distribution is positive, is a substantive implication of the interventionist account of causation (and the causal Markov condition). Even if Hausman and Woodward did not succeed in providing extra insights than what is usually offered to defend the Markov condition, anyone who actually performs causal inference based on the causal Markov condition has, in light of Theorem 1, an overwhelming reason to also assume the causal minimality condition, when the actual distribution is positive.

3 Causal minimality condition in the presence of determinism

The positivity assumption needed for Theorem 1 requires that for each variable X in $\mathbf{V}$, no value combination of its causal parents in $\mathbf{V}$ should completely determine what

[8] A perhaps better way to understand this is to take the conditional probabilities of a variable given its causal parents as primitives, which represent stable propensities of the local causal mechanism. And the problem of “undefined conditional probability” is the problem that in the actual population, some value setting of the causal parents may never be realized, and hence the propensity associated with that setting does not manifest in the actual population, and can’t be inferred from the actual distribution. To rigorously formulate this line would occupy much more space than allowed here.

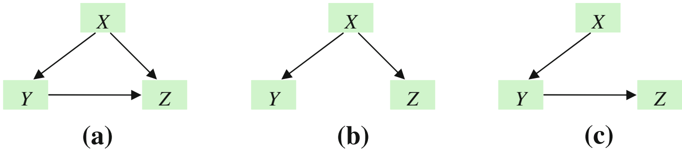

Fig. 1 A counterexample to the causal minimality condition. (**a**), referred to as graph G, is the true causal graph according to the mechanisms described in the text; (**b**) and (**c**), referred to as graph $G1$ and graph $G2$ respectively, are two proper subgraphs of (**a**) that satisfy the causal Markov condition

the value of X is, or what the value of X is *not*; that is, every possible value of X has a non-zero probability under every value setting of its causal parents. When some local mechanism is deterministic,[9] this assumption is immediately violated, and the causal minimality condition can fail.

Here is an extremely simple example. Suppose we have three switches, X, Y and Z, each of which is in one of two possible states, *on* (1) or *off* (0). The mechanisms are:

(a) X's state is determined by a toss of a fair coin: $P(X = 1) = P(X = 0) = 0.5$.
(b) Y is turned on iff X is turned on.
(c) Z is turned on with probability 0.8 if X and Y are both turned on; otherwise, Z is turned on with probability 0.2.

Notice that the local mechanism for Y is deterministic. Given this description, the causal structure in accord with the IIDC is obviously represented by the graph G in Fig. 1a. The joint probability distribution is also easy to calculate:

$$P(X = 1, Y = 1, Z = 1) = 0.4, \quad P(X = 1, Y = 1, Z = 0) = 0.1,$$
$$P(X = 0, Y = 0, Z = 1) = 0.1, \quad P(X = 0, Y = 0, Z = 0) = 0.4,$$

and other value settings have zero probability.

The causal minimality condition fails. It is easy to check that the actual distribution is also Markov to $G1$ (Fig. 1b) and to $G2$ (Fig. 1c), both of which are proper subgraphs of G.[10] Moreover, besides the presence of a deterministic mechanism, there is nothing special about this example, and it does not take long to convince oneself that when there are deterministic relationships, the failure of the condition is a rule rather than an exception.[11]

Are we then necessarily guilty if we make use of the causal minimality condition in causal inference? Let us pay a closer look at the two incorrect structures.

[9] Again, by "local mechanism" we mean that composed of the variables in $\mathbf{V}$. So X might be (objectively) governed by a deterministic mechanism, but if some of its causes are left out of $\mathbf{V}$, the local mechanism for X (relative to the variables in $\mathbf{V}$) may still be indeterministic.

[10] $G1$ entails that Y and Z are independent conditional on X. $G2$ entails that X and Z are independent conditional on Y. Both are true of the actual distribution.

[11] We should note that the failure of positivity does not require deterministic causal mechanisms in the system. It could just be that for some variables in the system, some combination of values has zero probability, for whatever reason. And the failure of positivity does not entail the failure of the causal minimality condition. What we claim here is simply that the causal minimality condition typically fails when positivity fails as a result of the presence of deterministic relationships. We thank a referee for pressing this point.

Take $G1$ for example. Suppose we mistake $G1$ for the true causal structure, what is the consequence for counterfactual reasoning, that is, for calculating post-intervention probabilities from the actual probability distribution? Here are some sample calculations according to the IP, taking $G1$ as the true causal structure:

$$P_{X:=1}(Y=1, Z=1) = P(Y=1|X=1)P(Z=1|X=1) = 0.8,$$
$$P_{Y:=0}(X=0, Z=1) = P(X=0)P(Z=1|X=0) = 0.1,$$
$$P_{Z:=1,X:=1}(Y=0) = P(Y=0|X=1) = 0.$$

All of them, one can easily check, are correct answers (calculated according to the true causal structure). This is also the case if we calculate these quantities using $G2$. Despite the simplicity of the examples, the correctness reflects a general fact:

Theorem 2 *Suppose G is the true causal structure for $\mathbf{V}$, and the IP holds. Let G' be any subgraph of G that is also Markov to the actual distribution. Then for any post-intervention probability $P_{\mathbf{S}:=\mathbf{s}}(\mathbf{T}=\mathbf{t})(\mathbf{S}, \mathbf{T} \subset \mathbf{V})$, if it can be inferred from the actual distribution based on G,*[12] *it can be calculated correctly based on G'.*

Proof Consider any intervention $do(\mathbf{S}=\mathbf{s})$. By the IP,

$$P_{\mathbf{S}:=\mathbf{s}}(\mathbf{V}\backslash\mathbf{S}) = \prod\nolimits_{X\in\mathbf{V}\backslash\mathbf{S}} P(\mathrm{X}|\mathbf{Pa}_G(\mathrm{X})),$$

whenever the conditional probabilities on the right hand side are defined. On the other hand, if we calculated $P_{\mathrm{s}}(\mathbf{V}\backslash\mathbf{S})$ based on G', we would get

$$P_{\mathbf{S}:=\mathbf{s}}(\mathbf{V}\backslash\mathbf{S}) = \prod\nolimits_{X\in\mathbf{V}\backslash\mathbf{S}} P(\mathrm{X}|\mathbf{Pa}_{G'}(X)).$$

So it suffices to show that for every X in $\mathbf{V}$, $P(\mathrm{X}|\mathbf{Pa}_{G'}(\mathrm{X})) = P(\mathrm{X}|\mathbf{Pa}_G(X))$, whenever the latter is defined. But this directly follows from the fact that $\mathbf{Pa}_{G'}(\mathrm{X}) \subseteq \mathbf{Pa}_G(\mathrm{X})$—because G' is a subgraph of G—and the fact that P is Markov to G' (and so X is independent of $\mathbf{Pa}_G(\mathrm{X})\backslash\mathbf{Pa}_{G'}(\mathrm{X})$ conditional on $\mathbf{Pa}_{G'}(\mathrm{X})$). □

Therefore, whatever counterfactual probability can be calculated from the actual probability distribution according to the true causal structure, can also be calculated *correctly* according to any substructure of the true one as long as the substructure is still Markov to the actual distribution.

This does not mean, however, that the incorrect causal structure does not get any counterfactual probability wrong. According to $G1$, for example,

$$P_{X:=1,Y:=0}(Z=1) = P(Z=1|X=1) = 0.8.$$

This is obviously far off given the specified mechanism for $Z: P_{X:=1,Y:=0}(Z=1) = 0.2$. In this case, the true causal structure fares better, not by giving the right answer,

[12] That is, when all the relevant conditional probabilities according to G are defined.

but by signaling that this quantity is not calculable from the actual distribution, because $P(Z|X = 1, Y = 0)$ is not defined.

Thus the incorrect causal structure may overshoot. Can we control that? It is of course no point recommending that one should only calculate those counterfactual quantities that can be calculated based on the true causal structure. Is there any criterion one can use that does not depend on knowing the true causal structure? Here is one such criterion.

Corollary 3 *Suppose the IP holds. Let G' be a substructure of the true causal structure that is Markov to the actual distribution. For any $\mathbf{S} \cup \mathbf{T} = \mathbf{V}$, if $P(\mathbf{S} = \mathbf{s}, \mathbf{T} = \mathbf{t}) > 0$, then the post-intervention probability $P_{\mathbf{S}:=\mathbf{s}}(\mathbf{T} = \mathbf{t})$ can be calculated correctly from the actual distribution based on G'.*

The corollary follows immediately from Theorem 2—because when $P(\mathbf{S}{=}\mathbf{s}, \mathbf{T}{=}\mathbf{t}) > 0$, every conditional probability needed to calculate $P_{\mathbf{S}:=\mathbf{s}}(\mathbf{T} = \mathbf{t})$ according to G is defined. Why is all this relevant to the causal minimality condition? Suppose the condition fails. That means some proper substructure of the true causal structure also satisfies the Markov condition. Then there exists a minimal substructure of the true causal structure that satisfies the Markov condition with the actual distribution. Consider any such minimal substructure. The foregoing discussion is supposed to show that this minimal substructure would still be useful for counterfactual reasoning. Corollary 3, in particular, gives us a manageable criterion for judging when we can use this object to do correct counterfactual reasoning. The criterion we expect can be further improved,[13] but that is a separate technical issue. The important epistemological point is that, for the purpose of counterfactual reasoning, the minimal substructure in question is a respectable target for causal inference.

With respect to this target, there is nothing wrong to assume the causal minimality condition. Hence, even in the presence of determinism, the causal minimality condition would be a harmless assumption, as long as we keep in mind the limitation of the targeted causal structure in aiding our counterfactual reasoning, and such criteria as the one stated in Corollary 3 that can guide our legitimate use of what is inferred based on the causal minimality condition.

Therefore, even when the causal minimality condition is literally false, the intervention principle entails a sort of methodological or pragmatic justification of the condition. This justification becomes especially compelling when the actual probability distribution is strictly positive. In that case, the sufficient condition for correctness given in Corollary 3 is always satisfied, which entails that, in terms of counterfactual reasoning, nothing is lost by appealing to a minimal substructure of the true causal structure in the sense that all post-intervention probabilities can be correctly calculated.

[13] Consider the sample calculations on page 10. For $P_{X:=1}(Y = 1, Z = 1)$ and $P_{Y:=0}(X = 0, Z = 1)$ the criterion is applicable, but the criterion does not authorize calculating $P_{Z:=1,X:=1}(Y = 0)$, even though the latter can be calculated correctly. We are searching for more powerful criteria. In the best scenario, the criterion could tell us exactly whether a counterfactual quantity is calculable from the (unknown) true causal structure. If so, using a minimal structure would entail no loss in power for calculating counterfactual distribution from the actual distribution. We are not optimistic about finding such a criterion, but we can't prove it is impossible at this point.

It is worth noting that, when positivity holds, this pragmatic justification is not entirely redundant given what we have shown in Sect. 2. The subtlety is that the pragmatic justification does not depend on the interventionist interpretation of causal arrows—the proof of Theorem 2 does not make use of IIDC. All it depends on is a conception of causal structure in terms of its epistemic role defined by the intervention principle.

To bring the last point home, the following result helps:

Theorem 4 *Let* $\mathbf{V}$ *be a set of variables. Suppose its actual joint distribution P is positive and some (acyclic) structure over* $\mathbf{V}$ *satisfies the IP. Then the structure as defined according to IIDC is the uniquely minimal structure that satisfies the IP.*

Proof Let G_{d} denote the structure defined by IIDC. We first show that any structure G that satisfies the IP is a superstructure of G_{d}. Suppose, for the sake of contradiction, G is not a superstructure of G_{d}. Then there is an arrow $X \rightarrow Y$ in G_{d} but not in G. Since G_{d} is defined according to IIDC, the presence of $X \rightarrow Y$ in G_{d} means that there exist $\mathrm{x} \neq \mathrm{x}'$ and $\mathbf{z}$ such that

$$P_{X:=x,\mathbf{Z}:=\mathbf{z}}(\mathrm{Y}) \neq P_{X:=x',\mathbf{Z}:=\mathbf{z}}(\mathrm{Y}), \quad \text{where } \mathbf{Z} = \mathbf{V}\backslash\{\mathrm{X}, \mathrm{Y}\}.$$

But G satisfies the IP, which means that

$$P_{X:=x,\mathbf{Z}:=\mathbf{z}}(\mathrm{Y}) = P_{X:=x',\mathbf{Z}:=\mathbf{z}}(\mathrm{Y}) = P(Y|\mathbf{Pa}_G(Y))$$

Because $X \notin \mathbf{Pa}_G(\mathrm{Y})$ (as $X \rightarrow Y$ is not in G). Hence we have a contradiction. So the original supposition is false, and G is a superstructure of G_{d}.

Next, we show that G_{d} satisfies the IP (if any structure does). Let G be a minimal structure that satisfies the IP, that is, a structure that satisfies the IP such that no proper substructure of G satisfies the IP. From the previous argument it follows that G is a superstructure of G_{d}. We now show that $G = G_{\mathrm{d}}$. Suppose not. Then G is a proper superstructure of G_{d}. So there is an edge $X \rightarrow Y$ in G which is not in G_{d}. Let G' be the same structure as G except that the edge $X \rightarrow Y$ is removed. We claim that G' also satisfies the IP. To show this, it suffices to show that

$$\mathrm{P}(Y|\mathbf{Pa}_G(Y)) = \mathrm{P}(Y|\mathbf{Pa}_{G'}(Y), X) = \mathrm{P}(Y|\mathbf{Pa}_{G'}(Y))$$

because other vertices have the exact same parents in G' as they do in G. But we have that for any value pa for $\mathbf{Pa}_{G'}(Y)$, and any $x \neq x'$,

$$\begin{aligned}\mathrm{P}(Y|\mathrm{Pa}_{G'}(Y) = \mathrm{pa}, X = x) &= \mathrm{P}_{\mathbf{Pa}_{G'}(Y):=\mathrm{pa}, X:=x, \mathbf{Z}:=\mathbf{z}}(Y)\\ &= \mathrm{P}_{\mathbf{Pa}_{G'}(Y):=\mathrm{pa}, X:=x', \mathbf{Z}:=\mathbf{z}}(Y) = \mathrm{P}(Y|\mathbf{Pa}_{G'}(Y) = \mathrm{pa}, X = x')\end{aligned}$$

where $\mathbf{Z} = \mathbf{V}\backslash(\{X, Y\} \cup \mathbf{Pa}_{G'}(Y))$, and $\mathbf{z}$ is any value for $\mathbf{Z}$. This is true because otherwise X would be a direct cause of Y according to IIDC, and hence would be a parent of Y in G_d. It follows that $\mathrm{P}(Y|\mathbf{Pa}_G(Y)) = \mathrm{P}(Y|\mathbf{Pa}_{G'}(Y))$, which implies that G' also satisfies the IP. But G' is a proper substructure of G, which contradicts our choice of G. So the supposition is false, and $G = G_d$. □

In more plain words, Theorem 4 says that the structure defined by IIDC is the simplest structure that satisfies the intervention principle (if any acyclic structure satisfies the principle). So, if we view causal structure functionally as that which can be used for counterfactual reasoning according to the intervention principle, then the structure defined by IIDC stands out as the natural candidate for being the "true" causal structure. In other words, when the probability distribution is positive, there is an excellent pragmatic motivation for IIDC itself, and with it, the causal minimality condition.

4 Conclusion

To summarize, the causal minimality condition in the theory of causal Bayesian networks is very well motivated in an interventionist account of causation. The connection can be seen in two ways when the actual distribution is positive. On the one hand, it is a logical consequence of the IIDC and the IP, the latter of which is but a combination of the interventionist notion of modularity with the causal Markov condition. On the other hand, it follows from the IP alone that the causal minimality condition would be a completely harmless assumption if all we care is the correctness of counterfactual reasoning.

The latter line of argument can be extended to situations where the actual distribution is not strictly positive, as we showed in Sect. 3. Although the technical results there may be further improved, the philosophical message should be clear: even when the causal minimality condition fails, assuming it in the inference of causal structures would be fine, if we put necessary restrictions on the subsequent use of inferred causal structures for counterfactual reasoning.

This point is general. The epistemology of causation has two sides: *hunting causes and using them*, to borrow the title of a recent book (Cartwright 2007). In evaluating a method for hunting causal structure, or its underlying assumptions, it is important to take into account how the inferred structure will be subsequently used. Conversely, how the structure is learned is also crucially relevant to how it should be used. The risk of false assumptions in hunting causes may be controlled in using them.

Finally, we should note that although the causal minimality condition clearly adds something to the causal Markov condition, the exact power and value of the condition for causal inference is not completely understood. One thing that is clear is that the condition is much less powerful than the better known and logically stronger *causal faithfulness condition*. However, it has been established recently that assuming the causal minimality condition, a significant portion of the causal faithfulness condition is empirically testable (Zhang and Spirtes 2008). For this reason at least, it is practically valuable to provide independent justifications for the causal minimality condition, regardless of whether it is also appropriate to assume the causal faithfulness condition.

Acknowledgements We thank Clark Glymour, Christopher Hitchcock, James Woodward, and two anonymous referees for valuable comments. Versions of this paper were presented at the twenty-first philosophy of science association biennial meeting and the monthly Monday meetings of formal epistemology at University of Konstanz. We thank the audience, especially Michael Baumgartner, Franz Huber, and Wolfgang Spohn for helpful criticisms and discussions

References

Campbell, J. (2007). An interventionist approach to causation in psychology. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy and computation* (pp. 58–66). New York: Oxford University Press.

Cartwright, N. (2002). Against modularity, the causal Markov condition and any link between the two: Comments on Hausman and Woodward. *British Journal for the Philosophy of Science, 53*, 411–453.

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge: Cambridge University Press.

Hausman, D., & Woodward, J. (1999). Independence, invariance and the causal Markov condition. *British Journal for the Philosophy of Science, 50*, 521–583.

Hausman, D., & Woodward, J. (2004a). Modularity and the causal Markov condition: A restatement. *British Journal for the Philosophy of Science, 55*, 147–161.

Hausman, D., & Woodward, J. (2004b). Manipulation and the causal Markov condition. *Philosophy of Science, 71*, 846–856.

Hitchcock, C. (2002). Probabilistic causation. In E. Zalta (Ed.), *Stanford Encyclopedia of philosophy*. http://plato.stanford.edu/entries/causation-probabilistic/.

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

Spirtes, P., Glymour, C., & Scheines, R. (1993) Causation, prediction and search. New York: Springer (2000, 2nd ed., Cambridge, MA: MIT Press).

Steel, D. (2006a). Methodological individualism, explanation, and invariance. *Philosophy of the Social Sciences, 36*, 440–463.

Steel, D. (2006b). Comment on Hausman and Woodward on the causal Markov condition. *British Journal for the Philosophy of Science, 57*, 219–231.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Zhang, J., & Spirtes, P. (2008). Detection of unfaithfulness and robust causal inference. *Minds and Machines, 18*(2), 239–271.