How Occam's Razor Provides a Neat Definition of Direct Causation

Alexander Gebharter & Gerhard Schurz
Duesseldorf Center for Logic and Philosophy of Science
University of Duesseldorf
Universitaetsstrasse 1
40225 Duesseldorf, Germany

Abstract

In this paper we show that the application of Occam's razor to the theory of causal Bayes nets gives us a neat definition of direct causation. In particular we show that Occam's razor implies Woodward's (2003) definition of direct causation, provided suitable intervention variables exist and the causal Markov condition (CMC) is satisfied. We also show how Occam's razor can account for direct causal relationships Woodward style when only stochastic intervention variables are available.

1 INTRODUCTION

Occam's razor is typically seen as a methodological principle. There are many possible ways to apply the razor to the theory of causal Bayes nets. It could, for example, simply be interpreted to suggest preferring the simplest causal structure compatible with the given data among all compatible causal structures. The simplest causal structure could, for instance, be the one (or one of the ones) featuring the fewest causal arrows.

In this paper, however, we are interested in a slightly different application of Occam's razor: Our interpretation of Occam's razor asserts that given a causal structure is compatible with the data, it should only be chosen if it satisfies the causal minimality condition (Min) in the sense of Spirtes et al. (2000, p. 31), which requires that no causal arrow in the structure can be omitted in such a way that the resulting substructure would still be compatible with the data. When speaking of a causal structure being compatible with the data, we have a causal structure and a probability distribution satisfying the causal Markov condition (CMC) in mind. (For details, see sec. 5.) In the following, applying Occam's razor always means to assume that the causal minimality condition is satisfied.

In this paper we give a motivation for Occam's razor that goes beyond its merits as a methodological principle dictating that one should always decide in favor of minimal causal models. In particular, we show that Occam's razor provides a neat definition of direct causal relatedness in the sense of Woodward (2003), provided suitable intervention variables exist and CMC is satisfied. Note the connection of this enterprise to Zhang and Spirtes' (2011) project. Zhang and Spirtes prove that CMC and an interventionist definition of direct causation a la Woodward (2003) together imply minimality. So Occam's razor is well-motivated within a manipulationist framework such as Woodward's. We show, vice versa, that CMC and minimality together imply Woodward's definition of direct causation. So if one wants a neat definition of direct causation, it is reasonable to apply Occam's razor in the sense of assuming minimality.

The paper is structured as follows: In sec. 2 we introduce the notation we use in subsequent sections. In sec. 3 we present Woodward's (2003) definition of direct causation and his definition of an intervention variable. In sec. 4 we give precise reconstructions of both definitions in terms of causal Bayes nets. We also provide a definition of the notion of an intervention expansion, which is needed to account for direct causal relations in terms of the existence of certain intervention variables. In sec. 5 we show that Occam's razor gives us Woodward's definition of direct causation if CMC is assumed and the existence of suitable intervention variables is granted (theorem 2). In sec. 6 we go a step further and show how Occam's razor allows us to account for direct causation Woodward style when only stochastic intervention variables (cf. Korb et al., 2004, sec. 5) are available (theorem 3). We conclude in sec. 7.

Note that though the main results of the present paper (i.e., theorems 2 and 3) can be used for causal discovery, the goal of this paper is not to provide a method for uncovering direct causal connections among variables in a set of variables $\mathbf{V}$ of interest. The goal of this paper is to establish a connection between Woodward's (2003) intervention-based notion of direct causation and the presence of a causal arrow in a minimal causal Bayes net, which can be interpreted as support for Occam's razor. Because of this, the present paper does not discuss the relation of theorems 2 and 3 to results about causal discovery by means of interventions such as, e.g., (Eberhardt and Scheines, 2007) or (Nyberg and Korb, 2007).

2 NOTATION

We represent causal structures by graphs, i.e., by ordered pairs $\langle \mathbf{V}, E\rangle$, where $\mathbf{V}$ is a set of variables and E is a binary relation on $\mathbf{V}$ ($E \subseteq \mathbf{V} \times \mathbf{V}$). $\mathbf{V}$'s elements are called the graph's "vertices" and E's elements are called its "edges". "$X \rightarrow Y$" stands short for "$\langle X, Y\rangle \in E$" and is interpreted as "X is a direct cause of Y in $\langle \mathbf{V}, E\rangle$" or as "$Y$ is a direct effect of X in $\langle \mathbf{V}, E\rangle$". $Par(Y)$ is the set of all $X \in \mathbf{V}$ with $X \rightarrow Y$ in $\langle \mathbf{V}, E\rangle$. The elements of $Par(Y)$ are called Y's parents. We write "$X - Y$" for "$X \rightarrow Y$ or $X \leftarrow Y$". A path $\pi : X - ... - Y$ is called a (causal) path connecting X and Y in $\langle \mathbf{V}, E\rangle$. A causal path π is called a directed causal path from X to Y if and only if ("iff" for short) it has the form $X \rightarrow ... \rightarrow Y$. X is called a cause of Y and Y an effect of X in that case. A causal path π is called a common cause path iff it has the form $X \leftarrow ... \leftarrow Z \rightarrow ... \rightarrow Y$ and no variable appears more often than once on π. Z is called a common cause of X and Y lying on path π in that case. A variable Z lying on a path $\pi : X - ... \rightarrow Z \leftarrow ... - Y$ is called a collider lying on this path. A variable X is called exogenous iff no arrow is pointing at X; it is called endogenous otherwise.

A graph $\langle \mathbf{V}, E\rangle$ is called a directed graph in case all edges in E are one-headed arrows "$\rightarrow$". It is called cyclic iff it features a causal path of the form $X \rightarrow ... \rightarrow X$ and acyclic otherwise. A causal structure $\langle \mathbf{V}, E\rangle$ together with a probability distribution P over $\mathbf{V}$ is called a causal model $\langle \mathbf{V}, E, P\rangle$. P is intended to provide information about the strengths of causal influences represented by the arrows in $\langle \mathbf{V}, E\rangle$. A causal model $\langle \mathbf{V}, E, P\rangle$ is called cyclic iff its graph $\langle \mathbf{V}, E\rangle$ is cyclic; it is called acyclic otherwise. In the following, we will only be interested in acyclic causal models.

We use the standard notions of (conditional) probabilistic dependence and independence:

Definition 1 (conditional probabilistic (in)dependence) *X and Y are probabilistically dependent conditional on Z iff there are X-, Y-, and Z-values x, y, and z, respectively, such that $P(x|y,z) \neq P(x|z) \wedge P(y,z) > 0$.*

X and Y are probabilistically independent conditional on Z iff X and Y are not probabilistically dependent conditional on Z.

Probabilistic independence between X and Y conditional on Z is abbreviated as "$Indep(X, Y|Z)$", probabilistic dependence is abbreviated as "$Dep(X, Y|Z)$". Unconditional probabilistic (in)dependence between X and Y $(In)Dep(X, Y)$ is defined as $(In)Dep(X, Y|\emptyset)$. X, Y, and Z in definition 1 can be variables or sequences of variables. When $X, Y, Z, ...$ are sequences of variables, we write them in bold letters. We write also the values $\mathbf{x}, \mathbf{y}, \mathbf{z}, ...$ of sequences $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, ...$ in bold letters. The set of values $\mathbf{x}$ of a sequence $\mathbf{X}$ of variables $X_1, ..., X_n$ is $val(X_1) \times ... \times val(X_n)$, where $val(X_i)$ is the set of X_i's possible values.

3 WOODWARD'S DEFINITION OF DIRECT CAUSATION

Woodward's (2003) interventionist theory of causation aims to explicate direct causation w.r.t. a set of variables $\mathbf{V}$ in terms of possible interventions. Woodward (2003, p. 98) provides the following definition of an intervention variable:

Definition 2 (IV$_W$) *I is an intervention variable for X with respect to Y if and only if I meets the following conditions:*
I1. I causes X.
I2. I acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.
I3. Any directed path from I to Y [if there exists one] goes through X [...].
I4. I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X.

(IV$_W$) is intended to single out those variables as intervention variables for X w.r.t. Y that allow for correct causal inference according to Woodward's (2003) definition of direct causation. For I to be an intervention variable for X w.r.t. Y it is required that I is causally relevant to X (condition I1), that X is only under I's influence when $I = on$ (condition I2), and that a correlation between I and Y can only be due to a directed causal path from I to Y going through X (conditions I3 and I4). For a detailed motivation of I1-I4, see (Woodward, 2003, sec. 3.1.4). For problems with Woodward's definitions, see (Gebharter and Schurz, ms).

An intervention on X w.r.t. Y (from now on we refer to X as the intervention's "target variable" and to Y as the "test variable") is then straightforwardly defined as an intervention variable I for X w.r.t. Y taking one of its on-values, which forces X to take a certain value x. We will call interventions whose on-values force X to take certain values x "deterministic interventions" (cf. Korb et al., 2004, sec. 5).

Note that Woodward's (2003) notion of an intervention is, on the one hand, strong because it requires interventions to be deterministic interventions. It is, on the other hand, weak in another respect: In contrast to structural or surgical interventions (cf. Eberhardt and Scheines, 2007, p. 984; Pearl, 2009) Woodward's interventions are allowed to be direct causes of more than one variable as long as the intervention's direct effects which are non-target variables do not cause the test variable over a path not going through the intervention's target variable (intervention condition I3).

Based on his notion of an intervention, Woodward (2003, p. 59) gives the following definition of direct causation w.r.t. a variable set $\mathbf{V}$:

Definition 3 (DC$_W$) *A necessary and sufficient condition for X to be a (type-level) direct cause of Y with respect to a variable set $\mathbf{V}$ is that there be a possible intervention on X that will change Y or the probability distribution of Y when one holds fixed at some value all other variables Z_i in $\mathbf{V}$.*

(DC$_W$) neatly explicates direct causation w.r.t. a variable set $\mathbf{V}$ in terms of possible interventions: X is a direct cause of Y w.r.t. $\mathbf{V}$ if Y can be wiggled by wiggling X; and if X is a direct cause of Y w.r.t. $\mathbf{V}$, then there are possible interventions by whose means one can influence Y by manipulating X.[1]

Note that **(DC$_W$)** may be too strong because many domains involve variables one cannot control by deterministic interventions. Scenarios of this kind include, for example, the decay of uranium or states of entangled systems in quantum mechanics. The decay of uranium can only be probabilistically influenced, and any attempt to manipulate the state of one of two entangled photons, for example, would destroy the entangled system. Glymour (2004) also considers variables for sex and race as not manipulable by means of intervention variables in the sense of **(IV$_W$)**.

To avoid all problems that might arise for Woodward's (2003) account due to variables that are not manipulable by deterministic interventions, we will reconstruct Woodward's **(DC$_W$)** as a partial definition in sec. 4. In particular, we will define direct causation only for sets of variables $\mathbf{V}$ for which suitable intervention variables exist.

4 RECONSTRUCTING WOODWARD'S DEFINITION

In this section we reconstruct Woodward's (2003) definition of direct causation in terms of causal Bayes nets. The reconstruction of **(IV$_W$)** is straightforward:

Definition 4 (IV) *$I_X \in \mathbf{V}$ is an intervention variable for $X \in \mathbf{V}$ w.r.t. $Y \in \mathbf{V}$ in a causal model $\langle \mathbf{V}, E, P \rangle$ iff*
(a) I_X is exogenous and there is a path $\pi : I_X \to X$ in $\langle \mathbf{V}, E \rangle$,
(b) for every on-value of I_X there is an X-value x such that $P(x|I_X = on) = 1$ and $Dep(x, I_X = on|\mathbf{z})$ holds for every instantiation $\mathbf{z}$ of every $\mathbf{Z} \subseteq \mathbf{V}\backslash\{I_X, X\}$,
(c) all paths $I_X \to ... \to Y$ in $\langle \mathbf{V}, E \rangle$ have the form $I_X \to ... \to X \to ... \to Y$,
(d) I_X is independent from every variable C (in $\mathbf{V}$ or not in $\mathbf{V}$) which causes Y over a path not going through X.

Note that **(IV)** still allows for intervention variables I_X that are common causes of their target variable X and other variables in $\mathbf{V}$. Condition (a) requires I_X to be exogenous. This is, though it is a typical assumption made for intervention variables, not explicit in Woodward's (2003) original definition **(IV$_W$)**. One problem that might arise for Woodward's account when not making this assumption is that I_X in a causal structure $Y \to I_X \to X$ may turn out to be an intervention variable for X w.r.t. Y. If Y then depends on $I_X = on$, **(DC$_W$)** would falsely determine X to be a cause of Y (cf. Gebharter and Schurz, ms). $I_X \to X$ in condition (a) is a harmless simplification of I1. Condition (b) captures Woodward's requirement that interventions have to be deterministic, from which I2 follows. X is assumed to be under full control of I_X when I_X is on. This does not only require that for every on-value of I_X there is an X-value x such that $P(x|I_X = on) = 1$, but also that $I_X = on$ actually has an influence on x in every possible context, i.e., under conditionalization on arbitrary instantiations $\mathbf{z}$ of all kinds of subsets $\mathbf{Z}$ of $\mathbf{V}\backslash\{I_X, X\}$. Condition (c) directly mirrors I3. Condition (d) mirrors Woodward's I4. Note that condition (d) requires reference to variables C possibly not contained in $\mathbf{V}$ (cf. Woodward, 2008, p. 202).

If we want to account for direct causal connection in a causal model $\langle \mathbf{V}, E, P \rangle$ by means of interventions, we have to add intervention variables to $\mathbf{V}$. In other words: We have to expand $\langle \mathbf{V}, E, P \rangle$ in a certain way. But how do we have to expand $\langle \mathbf{V}, E, P \rangle$? To answer this question, let us assume that we want to know whether X is a direct cause of Y in the unmanipulated model $\langle \mathbf{V}, E, P \rangle$. Then the manipulated model $\langle \mathbf{V}', E', P' \rangle$ will have to contain an intervention variable I_X for X w.r.t. Y and also intervention variables I_Z for all $Z \in \mathbf{V}$ different from X and Y by whose means these Z can be controlled. X is a direct cause of Y if I_X has some on-values such that we can influence Y by manipulating X with $I_X = on$ when all I_Z have taken certain on-values. On the other hand, to guarantee that X is not a direct cause of Y, we have to demonstrate that no one of Y's values can be influenced by manipulating some X-value by some intervention. For establishing such a negative causal claim, we require an intervention variable I_X by whose means we can control every X-value x. (Otherwise it could be that Y depends only on X-values that

[1]Note that Woodward (2003) does not require the intervention variables I to be elements of the set of variables $\mathbf{V}$ containing the target variable X and the test variable Y.

are not correlated with I_X-values; then $I_X = on$ would have no probabilistic influence on Y, though X may be a causal parent of Y.) In addition, we require for every $Z \neq X, Y$ an intervention variable I_Z by whose means Z can be forced to take every value z. (Otherwise it could be that we can bring about only such Z-value instantiations which screen X and Y off each other; then $I_X = on$ would have no probabilistic influence on Y when Z's value is fixed by interventions, though X may be a causal parent of Y.)

In the unmanipulated model $\langle \mathbf{V}, E, P\rangle$, all intervention variables I are off. In the manipulated model $\langle \mathbf{V}', E', P'\rangle$, all intervention variables' values are realized for some but not for all individuals in the domain. This move allows us to compute probabilities for variables in $\mathbf{V}$ when $I = off$ as well as probabilities for variables in $\mathbf{V}$ for all combinations of on-value realizations of intervention variables I, while the causal structure of the unmanipulated model will be preserved in the manipulated model. (Note that we deviate here from the typical "arrow breaking" representation of interventions in the literature which assumes that in the manipulated model all individuals get manipulated.) This amounts to the following notion of an intervention expansion ("i-expansion" for short):

Definition 5 (intervention expansion) *$\langle \mathbf{V}', E', P'\rangle$ is an intervention expansion of $\langle \mathbf{V}, E, P\rangle$ w.r.t. $Y \in \mathbf{V}$ iff*
(a) $\mathbf{V}' = \mathbf{V}\dot{\cup}\mathbf{V_I}$, where $\mathbf{V_I}$ contains for every $X \in \mathbf{V}$ different from Y an intervention variable I_X w.r.t. Y (and nothing else),
(b) for all $Z_i, Z_j \in \mathbf{V} : Z_i \rightarrow Z_j$ in E' iff $Z_i \rightarrow Z_j$ in E,
(c) for every X-value x of every $X \in \mathbf{V}$ different from Y there is an on-value of the corresponding intervention variable I_X such that $P'(x|I_X = on) = 1$ and $Dep(x, I_X = on|\mathbf{z})$ holds for every instantiation $\mathbf{z}$ of every $\mathbf{Z} \subseteq \mathbf{V}\backslash\{I_X, X\}$,
(d) $P'_{\mathbf{I=off}} \uparrow \mathbf{V} = P$,
(e) $P'(\mathbf{I = on}), P'(\mathbf{I = off}) > 0$.

$\mathbf{I}$ in conditions (d) and (e) is the set of all newly added intervention variables I. $P'_{\mathbf{I=off}} \uparrow \mathbf{V}$ in (d) is $P'_{\mathbf{I=off}} := P'(-|\mathbf{I = off})$ restricted to $\mathbf{V}$. Hence, "$P'_{\mathbf{I=off}} \uparrow \mathbf{V} = P$" means that $P'_{\mathbf{I=off}}$ coincides with P on the value space of variables in $\mathbf{V}$. Condition (a) guarantees that the i-expansion contains all the intervention variables required for testing for direct causal relationships in the sense of Woodward's (2003) definition of direct causation. The assumption that $\mathbf{V_I}$ contains only intervention variables for X w.r.t. Y is a harmless simplification. Thanks to condition (b), the manipulated model's causal structure fits to the unmanipulated model's causal structure. In particular, the i-expansion is only allowed to introduce new causal arrows going from intervention variables to variables in $\mathbf{V}$. Due to condition (c), every $X \in \mathbf{V}$ different from Y can be fully controlled by means of an intervention variable I_X for X w.r.t. Y. Condition (d) explains how the manipulated model's associated probability distribution P' fits to the unmanipulated model's distribution P. Condition (e) says that all values of intervention variables have to be realized by some individuals in the domain.

With help of this notion of an i-expansion we can now reconstruct Woodward's (2003) definition of direct causation. As already mentioned, Woodward's definition requires the existence of suitable intervention variables. Thus, we reconstruct **(DC$_W$)** as a partial definition whose if-condition presupposes the required intervention variables:

Definition 6 (DC) *If there exist i-expansions $\langle \mathbf{V}', E', P'\rangle$ of $\langle \mathbf{V}, E, P\rangle$ w.r.t. $Y \in \mathbf{V}$, then: $X \in \mathbf{V}$ is a direct cause of Y w.r.t. $\mathbf{V}$ iff $Dep(Y, I_X = on|\mathbf{I_Z} = \mathbf{on})$ holds in some i-expansions $\langle \mathbf{V}', E', P'\rangle$ of $\langle \mathbf{V}, E, P\rangle$ w.r.t. Y, where I_X is an intervention variable for X w.r.t. Y in $\langle \mathbf{V}', E', P'\rangle$ and $\mathbf{I_Z}$ is the set of all intervention variables in $\langle \mathbf{V}', E', P'\rangle$ different from I_X.*

(DC) mirrors Woodward's definition restricted to cases in which the required intervention variables (more precisely: the required i-expansions) exist: In case Y can be probabilistically influenced by manipulating X by means of an intervention variable I_X for X w.r.t. Y in one of these i-expansions, X is a direct cause of Y in the unmanipulated model. And vice versa: In case X is a direct cause of Y in the unmanipulated model, there will be an intervention variable I_X for X w.r.t. Y in one of these i-expansions such that Y is probabilistically sensitive to $I_X = on$.

In the next section we show that **(DC)** can account for all direct causal dependencies in a causal model if suitable i-expansions exist and CMC and Min are assumed to be satisfied.

5 OCCAM'S RAZOR, DETERMINISTIC INTERVENTIONS, AND DIRECT CAUSATION

The theory of causal Bayes nets' core axiom is the causal Markov condition (CMC) (cf. Spirtes et al., 2000, p. 29):

Definition 7 (causal Markov condition) *A causal model $\langle \mathbf{V}, E, P\rangle$ satisfies the causal Markov condition iff every $X \in \mathbf{V}$ is probabilistically independent of all its non-effects conditional on its causal parents.*

CMC is assumed to hold for causal models whose variable sets are causally sufficient. A variable set $\mathbf{V}$ is causally sufficient iff every common cause C of variables X and Y in $\mathbf{V}$ is also in $\mathbf{V}$ or takes the same value c for all individuals in the domain (cf. Spirtes et al., 2000, p. 22). From now on we implicitly assume causal sufficiency, i.e., we only consider causal models whose variable sets are causally sufficient.

A finite causal model $\langle \mathbf{V}, E, P \rangle$ satisfies the Markov condition iff P admits the following Markov factorization relative to $\langle \mathbf{V}, E \rangle$ (cf. Pearl, 2009, p. 16):

$$P(X_1, ..., X_n) = \prod_i P(X_i | Par(X_i)) \qquad (1)$$

The conditional probabilities $P(X_i | Par(X_i))$ are called X_i's parameters.

For acyclic causal models, CMC is equivalent to the d-separation criterion (Verma, 1986; Pearl, 1988, pp. 119f):

Definition 8 (d-separation criterion) *$\langle \mathbf{V}, E, P \rangle$ satisfies the d-separation criterion iff the following holds for all $X, Y \in \mathbf{V}$ and $\mathbf{Z} \subseteq \mathbf{V} \backslash \{X, Y\}$: If X and Y are d-separated by $\mathbf{Z}$ in $\langle \mathbf{V}, E \rangle$, then $Indep(X, Y | \mathbf{Z})$.*

Definition 9 (d-separation, d-connection) *$X \in \mathbf{V}$ and $Y \in \mathbf{V}$ are d-separated by $\mathbf{Z} \subseteq \mathbf{V} \backslash \{X, Y\}$ in $\langle \mathbf{V}, E \rangle$ iff X and Y are not d-connected given $\mathbf{Z}$ in $\langle \mathbf{V}, E \rangle$.*

$X \in \mathbf{V}$ and $Y \in \mathbf{V}$ are d-connected given $\mathbf{Z} \subseteq \mathbf{V} \backslash \{X, Y\}$ in $\langle \mathbf{V}, E \rangle$ iff X and Y are connected by a path π in $\langle \mathbf{V}, E \rangle$ such that no non-collider on π is in $\mathbf{Z}$, while all colliders on π are in $\mathbf{Z}$ or have an effect in $\mathbf{Z}$.

The equivalence between CMC and the d-separation criterion reveals the full content of CMC: If a causal model satisfies CMC, then every (conditional) probabilistic independence can be explained by missing (conditional) causal connections, and every (conditional) probabilistic dependence can be explained by some existing (conditional) causal connection.

In case there is a path π between X and Y in $\langle \mathbf{V}, E \rangle$ such that no non-collider on π is in $\mathbf{Z} \subseteq \mathbf{V} \backslash \{X, Y\}$ and all colliders on π are in $\mathbf{Z}$ or have an effect in $\mathbf{Z}$, π is said to be activated by $\mathbf{Z}$. We also say that X and Y are d-connected given $\mathbf{Z}$ over path π in that case. If π is not activated by $\mathbf{Z}$, π is said to be blocked by $\mathbf{Z}$. We also say that X and Y are d-separated by $\mathbf{Z}$ over path π in that case.

Occam's razor (as we understand it in this paper) dictates to prefer from all those causal structures $\langle \mathbf{V}, E \rangle$, which together with a given probability distribution P over $\mathbf{V}$ satisfy CMC, the ones which also satisfy the causal minimality condition (Min):

Definition 10 (causal minimality condition) *A causal model $\langle \mathbf{V}, E, P \rangle$ satisfying CMC satisfies the causal minimality condition iff no model $\langle \mathbf{V}, E', P \rangle$ with $E' \subset E$ also satisfies CMC (cf. Spirtes et al., 2000, p. 31).*

For acyclic causal models satisfying CMC, the following causal productivity condition (Prod) (cf. Schurz and Gebharter, forthcoming) can be seen as a reformulation of the causal minimality condition:

Definition 11 (causal productivity condition) *A causal model $\langle \mathbf{V}, E, P \rangle$ satisfies the causal productivity condition iff $Dep(X, Y | Par(Y) \backslash \{X\})$ holds for all $X, Y \in \mathbf{V}$ with $X \rightarrow Y$ in $\langle \mathbf{V}, E \rangle$.*

Theorem 1 *For every acyclic causal model $\langle \mathbf{V}, E, P \rangle$ satisfying CMC, the causal minimality condition and the causal productivity condition are equivalent.*

The equivalence of Min and Prod reveals the full content of Min: In minimal causal models, no causal arrow is superfluous, i.e., every causal arrow from X to Y is productive, meaning that it is responsible for some probabilistic dependence between X and Y (when the values of all other parents of Y are fixed).

We can now prove the following theorem:

Theorem 2 *If $\langle \mathbf{V}, E, P \rangle$ is an acyclic causal model and for every $Y \in \mathbf{V}$ there is an i-expansion $\langle \mathbf{V}', E', P' \rangle$ of $\langle \mathbf{V}, E, P \rangle$ w.r.t. Y satisfying CMC and Min, then for all $X, Y \in \mathbf{V}$ (with $X \neq Y$) the following two statements are equivalent:*
(i) $X \rightarrow Y$ in $\langle \mathbf{V}, E \rangle$.
(ii) $Dep(Y, I_X = on | \mathbf{I_Z} = \mathbf{on})$ holds in some i-expansions $\langle \mathbf{V}', E', P' \rangle$ of $\langle \mathbf{V}, E, P \rangle$ w.r.t. Y, where I_X is an intervention variable for X w.r.t. Y in $\langle \mathbf{V}', E', P' \rangle$ and $\mathbf{I_Z}$ is the set of all intervention variables in $\langle \mathbf{V}', E', P' \rangle$ different from I_X.

Theorem 2 shows that direct causation a la Woodward (2003) coincides with the graph theoretical notion of direct causation in systems $\langle \mathbf{V}, E, P \rangle$ with i-expansions w.r.t. every variable $Y \in \mathbf{V}$ satisfying CMC and Min. In particular, theorem 2 says the following: Assume we are interested in a causal model $\langle \mathbf{V}, E, P \rangle$. Assume further that for every Y in $\mathbf{V}$ there is an i-expansion $\langle \mathbf{V}', E', P' \rangle$ of $\langle \mathbf{V}, E, P \rangle$ w.r.t. Y satisfying CMC and Min. This means (among other things) that for every pair of variables $\langle X, Y \rangle$ there is at least one i-expansion with an intervention variable I_X for X w.r.t. Y and intervention variables I_Z for every $Z \in \mathbf{V}$ (different from X and Y) w.r.t. Y by whose means one can force the variables in $\mathbf{V} \backslash \{Y\}$ to take any combination of value realizations. Given this setup, theorem 2 tells us for every X and Y (with $X \neq Y$) in $\mathbf{V}$ that X is a causal parent of Y in $\langle \mathbf{V}, E \rangle$ iff $Dep(Y, I_X = on | \mathbf{I_Z} = \mathbf{on})$ holds in one of the presupposed i-expansions w.r.t. Y.

6 OCCAM'S RAZOR, STOCHASTIC INTERVENTIONS, AND DIRECT CAUSATION

In this section we generalize the main finding of sec. 5 to cases in which only stochastic interventions are available. To account for direct causal relations $X \rightarrow Y$ by means of stochastic intervention variables, two intervention variables are needed, one for X and one for Y. (For details, see below.) We define a stochastic intervention variable as follows:

Definition 12 ($\mathbf{IV}_S$) *$I_X \in \mathbf{V}$ is a stochastic intervention variable for $X \in \mathbf{V}$ w.r.t. $Y \in \mathbf{V}$ in $\langle \mathbf{V}, E, P \rangle$ iff*
(a) I_X is exogenous and there is a path $\pi : I_X \to X$ in $\langle \mathbf{V}, E \rangle$,
(b) for every on-value of I_X there is an X-value x such that $Dep(x, I_X = on|\mathbf{z})$ holds for every instantiation $\mathbf{z}$ of every $\mathbf{Z} \subseteq \mathbf{V} \backslash \{I_X, X\}$,
(c) all paths $I_X \to ... \to Y$ in $\langle \mathbf{V}, E \rangle$ have the form $I_X \to ... \to X \to ... \to Y$,
(d) I_X is independent from every variable C (in $\mathbf{V}$ or not in $\mathbf{V}$) which causes Y over a path not going through X.

The only difference between **($\mathbf{IV}_S$)** and **(IV)** is condition (b). For stochastic interventions it is not required that $I_X = on$ determines X's value to be x with probability 1. It suffices that $I_X = on$ and x are correlated conditional on every value $\mathbf{z}$ of every $\mathbf{Z} \subseteq \mathbf{V} \backslash \{I_X, X\}$. This specific constraint guarantees that X can be influenced by $I_X = on$ under all circumstances, i.e., under all kinds of conditionalization on instantiations of remainder variables in $\mathbf{V}$.

We do also have to modify our notion of an intervention expansion in case we allow for stochastic interventions. We define the following notion of a stochastic intervention expansion:

Definition 13 (stochastic intervention expansion) *$\langle \mathbf{V}', E', P' \rangle$ is a stochastic intervention expansion of $\langle \mathbf{V}, E, P \rangle$ for $X \in \mathbf{V}$ w.r.t. $Y \in \mathbf{V}$ iff*
(a) $\mathbf{V}' = \mathbf{V} \dot{\cup} \mathbf{V_I}$, where $\mathbf{V_I}$ contains one stochastic intervention variable I_X for X w.r.t. Y and one stochastic intervention variable I_Y for Y w.r.t. Y which is a parent only of Y (and nothing else),
(b) for all $Z_i, Z_j \in \mathbf{V} : Z_i \to Z_j$ in E' iff $Z_i \to Z_j$ in E,
(c.1) for every X-value x there is an on-value of I_X such that $Dep(x, I_X = on|\mathbf{z})$ holds for every instantiation $\mathbf{z}$ of every $\mathbf{Z} \subseteq \mathbf{V}' \backslash \{I_X, X\}$,
(c.2) for every Y-value y, every instantiation $\mathbf{r}$ of $Par(Y)$, and every on-value of I_Y there is an on-value on^* of I_Y such that $P'(y|I_Y = on^*, \mathbf{r}) \neq P'(y|I_Y = on, \mathbf{r})$, $P'(y|I_Y = on^*, \mathbf{r}) > 0$, and $P'(y|I_Y = on^*, \mathbf{r}^*) = P'(y|I_Y = on, \mathbf{r}^*)$ holds for all $\mathbf{r}^* \in val(Par(Y))$ different from $\mathbf{r}$,
(d) $P'_{\mathbf{I=off}} \uparrow \mathbf{V} = P$,
(e) $P'(\mathbf{I} = \mathbf{on}), P'(\mathbf{I} = \mathbf{off}) > 0$.

This definition differs from the definition of a (non-stochastic) i-expansion with respect to conditions (a) and (c): A stochastic i-expansion for X w.r.t. Y contains exactly two intervention variables, viz. one stochastic intervention variable I_X for X w.r.t. Y and one stochastic intervention variable I_Y for Y w.r.t. Y (which trivially satisfies conditions (c) and (d) in **($\mathbf{IV}_S$)**). While I_X may have more than one direct effect, the second intervention variable I_Y is assumed to be a causal parent only of Y. (This is required for accounting for direct causal connections; for details see (i) ⇒ (ii) in the proof of theorem 3 in the appendix.)

The second intervention variable I_Y is required to exclude independence between I_X and Y due to a fine-tuning of Y's parameters. Such an independence can arise even if CMC and Min are satisfied, X is a causal parent of Y, and I_X and Y are each correlated with the same X-values x. For examples of this kind of non-faithfulness, see, e.g., (Neapolitan, 2004, p. 96) or (Naeger, forthcoming). In condition (c.2) we assume that every one of Y's parameters can be changed independently of all other Y-parameters (to a value $r \in]0, 1]$) by changing I_Y's on-value. This suffices to exclude non-faithful independencies between I_X and Y of the kind described above.

When not presupposing deterministic interventions, it cannot be guaranteed anymore that the value of every variable in our model of interest different from the test variable Y can be fixed by interventions. The values of a causal model's variables can, however, also be fixed by conditionalization. To account for direct causation between X and Y when only stochastic interventions are available, one has to conditionalize on a suitably chosen set $\mathbf{Z} \subseteq \mathbf{V} \backslash \{X, Y\}$ that (i) blocks all indirect causal paths between X and Y, and that (ii) fixes all X-alternative parents of Y. That $\mathbf{Z}$ blocks all indirect paths between X and Y is required to assure that dependence between $I_X = on$ and Y cannot be due to an indirect path, and fixing the values of all parents of Y different from X is required to exclude independence of $I_X = on$ and Y due to a fine-tuning of Y's X-alternative parents that may cancel the influence of $I_X = on$ on Y over a path $I_X \to X \to Y$.[2] Fortunately, every directed acyclic graph $\langle \mathbf{V}, E \rangle$ features a set $\mathbf{Z}$ satisfying requirement (i), viz. $Par(Y) \backslash \{X\}$ (cf. Schurz and Gebharter, forthcoming). Trivially, $Par(Y) \backslash \{X\}$ also satisfies requirement (ii).

With the help of **($\mathbf{IV}_S$)** and definition 13, we can now define direct causation in terms of stochastic interventions for models for which suitable stochastic i-expansions exist:

Definition 14 ($\mathbf{DC}_S$) *If there exist stochastic i-expansions $\langle \mathbf{V}', E', P' \rangle$ of $\langle \mathbf{V}, E, P \rangle$ for X w.r.t. Y, then: X is a direct cause of Y w.r.t. $\mathbf{V}$ iff $Dep(Y, I_X = on|Par(Y) \backslash \{X\}, I_Y = on)$ holds in some i-expansions $\langle \mathbf{V}', E', P' \rangle$ of $\langle \mathbf{V}, E, P \rangle$ for X w.r.t. Y, where I_X is a stochastic intervention variable for X w.r.t. Y in $\langle \mathbf{V}', E', P' \rangle$ and I_Y is a stochastic intervention variable for Y w.r.t. Y in $\langle \mathbf{V}', E', P' \rangle$.*

Now the following theorem can be proven:

[2]For details on such cases of non-faithfulness due to compensating parents see (Schurz and Gebharter, forthcoming; Pearl, 1988, p. 256).

Theorem 3 *If $\langle \mathbf{V}, E, P\rangle$ is an acyclic causal model and for every $X, Y \in \mathbf{V}$ (with $X \neq Y$) there is a stochastic i-expansion $\langle \mathbf{V}', E', P'\rangle$ of $\langle \mathbf{V}, E, P\rangle$ for X w.r.t. Y satisfying CMC and Min, then for all $X, Y \in \mathbf{V}$ (with $X \neq Y$) the following two statements are equivalent:*
(i) $X \to Y$ in $\langle \mathbf{V}, E\rangle$.
(ii) $Dep(Y, I_X = on | Par(Y)\backslash\{X\}, I_Y = on)$ holds in some i-expansions $\langle \mathbf{V}', E', P'\rangle$ of $\langle \mathbf{V}, E, P\rangle$ for X w.r.t. Y, where I_X is a stochastic intervention variable for X w.r.t. Y in $\langle \mathbf{V}', E', P'\rangle$ and I_Y is a stochastic intervention variable for Y w.r.t. Y in $\langle \mathbf{V}', E', P'\rangle$.

Theorem 3 shows that direct causation a la Woodward (2003) coincides with the graph theoretical notion of direct causation in systems $\langle \mathbf{V}, E, P\rangle$ with stochastic i-expansions for every $X \in \mathbf{V}$ w.r.t. every $Y \in \mathbf{V}$ (with $X \neq Y$) satisfying CMC and Min. In particular, theorem 3 says the following: Assume we are interested in a causal model $\langle \mathbf{V}, E, P\rangle$. Assume further that for every X, Y in $\mathbf{V}$ (with $X \neq Y$) there is a stochastic i-expansion $\langle \mathbf{V}', E', P'\rangle$ of $\langle \mathbf{V}, E, P\rangle$ for X w.r.t. Y satisfying CMC and Min. This means (among other things) that for every pair of variables $\langle X, Y\rangle$ there is at least one stochastic i-expansion featuring a stochastic intervention variable I_X for X w.r.t. Y and a stochastic intervention variable I_Y for Y w.r.t. Y. Given this setup, theorem 3 can account for every causal arrow between every X and Y (with $X \neq Y$) in $\mathbf{V}$: It says that X is a causal parent of Y in $\langle \mathbf{V}, E\rangle$ iff $Dep(Y, I_X = on | Par(Y)\backslash\{X\}, I_Y = on)$ holds in some of the presupposed stochastic i-expansions for X w.r.t. Y.

7 CONCLUSION

In this paper we investigated the consequences of assuming a certain version of Occam's razor. If one applies the razor in such a way to the theory of causal Bayes nets that it dictates to prefer only minimal causal models, one can show that Occam's razor provides a neat definition of direct causation. In particular, we demonstrated that one gets Woodward's (2003) definition of direct causation translated into causal Bayes nets terminology and restricted to contexts in which suitable i-expansions satisfying the causal Markov condition (CMC) exist. In the last section we showed how Occam's razor can be used to account for direct causal connections Woodward style even if no deterministic interventions are available. These results can be seen as a motivation of Occam's razor going beyond its merits as a methodological principle: If one wants a nice and simple interventionist definition of direct causation in the sense of Woodward (or its stochastic counterpart developed in sec. 6), then it is reasonable to apply a version of Occam's razor that suggests to eliminate non-minimal causal models.

Acknowledgements

This work was supported by DFG, research unit "Causation, Laws, Dispositions, Explanation" (FOR 1063). Our thanks go to Frederick Eberhardt and Paul Naeger for important discussions, to two anonymous referees for helpful comments on an earlier version of the paper, and to Sebastian Maaß for proofreading.

References

F. Eberhardt, and R. Scheines (2007). Interventions and causal inference. *Philosophy of Science* **74**(5):981-995.

A. Gebharter, and G. Schurz (ms). Woodward's interventionist theory of causation: Problems and proposed solutions.

C. Glymour (2004). Critical notice. *British Journal for the Philosophy of Science* **55**(4):779-790.

K. B. Korb, L. R. Hope, A. E. Nicholson, and K. Axnick (2004). Varieties of causal intervention. In C. Zhang, H. W. Guesgen, W.-K. Yeap (eds.), *Proceedings of the 8th Pacific Rim International Conference on AI 2004: Trends in Artificial Intelligence*, 322-331. Berlin: Springer.

P. Naeger (forthcoming). The causal problem of entanglement. *Synthese*.

R. Neapolitan (2004). *Learning Bayesian Networks*. Upper Saddle River, NJ: Prentice Hall.

E. P. Nyberg, and K. B. Korb (2006). Informative interventions. Technical report 2006/204, Clayton School of Information Technology, Monash University, Melbourne.

J. Pearl (1988). *Probabilistic Reasoning in Expert Systems*. San Mateo, MA: Morgan Kaufmann.

J. Pearl (2009). *Causality*. Cambridge: Cambridge University Press.

G. Schurz, and A. Gebharter (forthcoming). Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets. *Synthese*.

P. Spirtes, C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search*. Cambridge, MA: MIT Press.

T. S. Verma (1986). Causal networks: Semantics and expressiveness. Technical report R-65, Cognitive Systems Laboratory, University of California, Los Angeles.

J. Woodward (2003). *Making Things Happen*. Oxford: Oxford University Press.

J. Woodward (2008). Response to Strevens. *Philosophy and Phenomenological Research* **77**(1):193-212.

J. Zhang, and P. Spirtes (2011). Intervention, determinism, and the causal minimality condition. *Synthese* **182**(3):335-347.

Appendix

The following proof of theorem 1 rests on the equivalence of CMC and the Markov factorization (1). It is, thus, restricted to finite causal structures.

Proof of theorem 1 Suppose $\langle \mathbf{V}, E, P\rangle$ with $\mathbf{V} = \{X_1, ..., X_n\}$ to be a finite acyclic causal model satisfying CMC.

Prod ⇒ Min: Assume that $\langle \mathbf{V}, E, P\rangle$ does not satisfy Min, meaning that there are $X, Y \in \mathbf{V}$ with $X \to Y$ in $\langle \mathbf{V}, E\rangle$ such that $\langle \mathbf{V}, E', P\rangle$, which results from deleting $X \to Y$ from $\langle \mathbf{V}, E\rangle$, still satisfies CMC. But then $Par(Y)\backslash\{X\}$ d-separates X and Y in $\langle \mathbf{V}, E'\rangle$, and thus, the d-separation criterion implies $Indep(X, Y|Par(Y)\backslash\{X\})$, which violates Prod.

Min ⇒ Prod: Assume that $\langle \mathbf{V}, E, P\rangle$ satisfies Min, meaning that there are no $X, Y \in \mathbf{V}$ with $X \to Y$ in $\langle \mathbf{V}, E\rangle$ such that $\langle \mathbf{V}, E', P\rangle$, which results from deleting $X \to Y$ from $\langle \mathbf{V}, E\rangle$, still satisfies CMC. The latter is the case iff (*) the parent set $Par(Y)$ of every $Y \in \mathbf{V}$ (with $Par(Y) \neq \emptyset$) is minimal in the sense that removing one of Y's parents X from $Par(Y)$ would make a difference for Y, meaning that $P(y|x, Par(Y)\backslash\{X\} = \mathbf{r}) \neq P(y|Par(Y)\backslash\{X\} = \mathbf{r})$ holds for some X-values x, some Y-values y, and some instantiations $\mathbf{r}$ of $Par(Y)\backslash\{X\}$. Otherwise P would admit the Markov factorization relative to $\langle \mathbf{V}, E\rangle$ and relative to $\langle \mathbf{V}, E'\rangle$, meaning that also $\langle \mathbf{V}, E', P\rangle$, which results from deleting $X \to Y$ from $\langle \mathbf{V}, E\rangle$, would satisfy CMC. But then $\langle \mathbf{V}, E, P\rangle$ would not be minimal, which would contradict the assumption. Now (*) entails that $Dep(X, Y|Par(Y)\backslash\{X\})$ holds for all $X, Y \in \mathbf{V}$ with $X \to Y$, i.e., that $\langle \mathbf{V}, E, P\rangle$ satisfies Prod. □

Proof of theorem 2 Assume $\langle \mathbf{V}, E, P\rangle$ is an acyclic causal model and for every $Y \in \mathbf{V}$ there is an i-expansion $\langle \mathbf{V}', E', P'\rangle$ of $\langle \mathbf{V}, E, P\rangle$ w.r.t. Y satisfying CMC and Min. Let X and Y be arbitrarily chosen elements of $\mathbf{V}$ such that $X \neq Y$.

(i) ⇒ (ii): Suppose $X \to Y$ in $\langle \mathbf{V}, E\rangle$. We assumed that there exists an i-expansion $\langle \mathbf{V}', E', P'\rangle$ of $\langle \mathbf{V}, E, P\rangle$ w.r.t. Y satisfying CMC and Min. From condition (b) of definition 5 it follows that $X \to Y$ in $\langle \mathbf{V}', E'\rangle$. Since Min is equivalent to Prod, X and Y are dependent when the values of all parents of Y different from X are fixed to certain values, meaning that there will be an X-value x and a Y-value y such that $Dep(x, y|Par(Y)\backslash\{X\} = \mathbf{r})$ holds for an instantiation $\mathbf{r}$ of $Par(Y)\backslash\{X\}$. Now there will also be a value of $\mathbf{I_Z}$ that fixes the set of all parents of Y different from X to $\mathbf{r}$. Let $\mathbf{on}$ be this $\mathbf{I_Z}$-value. Thus, also $Dep(x, y|\mathbf{I_Z} = \mathbf{on})$ and also $Dep(x, y|\mathbf{I_Z} = \mathbf{on}, \mathbf{r})$ will hold. Now let us assume that on is one of the I_X-values which are correlated with x and which force X to take value x. (The existence of such an I_X-value is guaranteed by condition (c) in definition 5.) Then we have $Dep(I_X = on, x|\mathbf{I_Z} = \mathbf{on}, \mathbf{r}) \wedge Dep(x, y|\mathbf{I_Z} = \mathbf{on}, \mathbf{r})$. From the axiom of weak union (2) (cf. Pearl, 2009, p. 11), which is probabilistically valid, we get (3) and (4) (in which $\mathbf{s} = \langle x, \mathbf{r}\rangle$ is a value realization of $Par(Y)$):

$$Indep(X, YW|Z) \Rightarrow Indep(X, Y|ZW) \tag{2}$$

$$\begin{aligned} &Indep(I_X = on, \mathbf{s} = \langle x, \mathbf{r}\rangle|\mathbf{I_Z} = \mathbf{on}) \Rightarrow \\ &Indep(I_X = on, x|\mathbf{I_Z} = \mathbf{on}, \mathbf{r}) \end{aligned} \tag{3}$$

$$\begin{aligned} &Indep(\mathbf{s} = \langle x, \mathbf{r}\rangle, y|\mathbf{I_Z} = \mathbf{on}) \Rightarrow \\ &Indep(x, y|\mathbf{I_Z} = \mathbf{on}, \mathbf{r}) \end{aligned} \tag{4}$$

With the contrapositions of (3) and (4) it now follows that $Dep(I_X = on, \mathbf{s} = \langle x, \mathbf{r}\rangle|\mathbf{I_Z} = \mathbf{on}) \wedge Dep(\mathbf{s} = \langle x, \mathbf{r}\rangle, y|\mathbf{I_Z} = \mathbf{on})$.

We now show that $Dep(I_X = on, \mathbf{s}|\mathbf{I_Z} = \mathbf{on}) \wedge Dep(\mathbf{s}, y|\mathbf{I_Z} = \mathbf{on})$ and the d-separation criterion imply $Dep(I_X = on, y|\mathbf{I_Z} = \mathbf{on})$. We define $P^*(-)$ as $P'(-|\mathbf{I_Z} = \mathbf{on})$ and proceed as follows:

$$\begin{aligned} &P^*(y|I_X = on) = \\ &\sum_i P^*(y|\mathbf{s}_i, I_X = on) \cdot P^*(\mathbf{s}_i|I_X = on) \end{aligned} \tag{5}$$

Equation (5) is probabilistically valid. Because $Par(Y)$ blocks all paths between I_X and Y, we get (6) from (5):

$$\begin{aligned} &P^*(y|I_X = on) = \\ &\sum_i P^*(y|\mathbf{s}_i) \cdot P^*(\mathbf{s}_i|I_X = on) \end{aligned} \tag{6}$$

Since $I_X = on$ forces $Par(Y)$ to take value $\mathbf{s}$ when $\mathbf{I_Z} = \mathbf{on}$, $P^*(\mathbf{s}_i|I_X = on) = 1$ in case $\mathbf{s}_i = \mathbf{s}$, and $P^*(\mathbf{s}_i|I_X = on) = 0$ otherwise. Thus, we get (7) from (6):

$$P^*(y|I_X = on) = P^*(y|\mathbf{s}) \cdot 1 \tag{7}$$

For reductio, let us assume that $Indep(I_X = on, y|\mathbf{I_Z} = \mathbf{on})$, meaning that $P^*(y|I_X = on) = P^*(y)$. But then we get (8) from (7):

$$P^*(y) = P^*(y|\mathbf{s}) \cdot 1 \tag{8}$$

Equation (8) contradicts $Dep(\mathbf{s}, y|\mathbf{I_Z} = \mathbf{on})$ above. Hence, $Dep(I_X = on, y|\mathbf{I_Z} = \mathbf{on})$ has to hold when $Dep(I_X = on, \mathbf{s}|\mathbf{I_Z} = \mathbf{on}) \wedge Dep(\mathbf{s}, y|\mathbf{I_Z} = \mathbf{on})$ holds. Therefore, $Dep(Y, I_X = on|\mathbf{I_Z} = \mathbf{on})$.

(ii) ⇒ (i): Suppose $\langle \mathbf{V}', E', P'\rangle$ is one of the presupposed i-expansions such that $Dep(Y, I_X = on|\mathbf{I_Z} = \mathbf{on})$ holds, where I_X is an intervention variable for X w.r.t. Y in $\langle \mathbf{V}', E', P'\rangle$ and $\mathbf{I_Z}$ is the set of all intervention variables in $\langle \mathbf{V}', E', P'\rangle$ different from I_X. Then the d-separation criterion implies that there must be a causal path π d-connecting I_X and Y. π cannot be a path featuring colliders, because I_X and Y would be d-separated over such

a path. π also cannot have the form $I_X \leftarrow ... - Y$. This is excluded by condition (a) in **(IV)**. So π must have the form $I_X \rightarrow ... - Y$. Since π cannot feature colliders, π must be a directed path $I_X \rightarrow ... \rightarrow Y$. Now either (A) π goes through X, or (B) π does not go through X. (B) is excluded by condition (c) in **(IV)**. Hence, (A) must be the case. If (A) is the case, then π is a directed path $I_X \rightarrow ... \rightarrow X \rightarrow ... \rightarrow Y$ going through X. Now there are two possible cases: Either (i) at least one of the paths π d-connecting I_X and Y has the form $I_X \rightarrow ... \rightarrow X \rightarrow Y$, or (ii) all paths π d-connecting I_X and Y have the form $I_X \rightarrow ... \rightarrow X \rightarrow ... \rightarrow C \rightarrow ... \rightarrow Y$.

Assume (ii) is the case, i.e., all paths π d-connecting I_X and Y have the form $I_X \rightarrow ... \rightarrow X \rightarrow ... \rightarrow C \rightarrow ... \rightarrow Y$. Let $\mathbf{r}_i$ be an individual variable ranging over $val(Par(Y))$. We define $P^*(-)$ as $P'(-|\mathbf{I_Z} = \mathbf{on})$ and proceed as follows:

$$P^*(y|I_X = on) = \sum_i P^*(y|\mathbf{r}_i, I_X = on) \cdot P^*(\mathbf{r}_i|I_X = on) \tag{9}$$

$$P^*(y) = \sum_i P^*(y|\mathbf{r}_i) \cdot P^*(\mathbf{r}_i) \tag{10}$$

Equations (9) and (10) are probabilistically valid. Since $\mathbf{I_Z} = \mathbf{on}$ forces every non-intervention variable in $\mathbf{V}'$ different from X and Y to take a certain value, $\mathbf{I_Z} = \mathbf{on}$ will also force $Par(Y)$ to take a certain value $\mathbf{r}$, meaning that $P^*(\mathbf{r}_i) = 1$ in case $\mathbf{r}_i = \mathbf{r}$, and that $P^*(\mathbf{r}_i) = 0$ otherwise. Since probabilities of 1 do not change after conditionalization, we get $P^*(\mathbf{r}_i|I_X = on) = 1$ in case $\mathbf{r}_i = \mathbf{r}$, and $P^*(\mathbf{r}_i|I_X = on) = 0$ otherwise. Thus, we get (11) from (9) and (12) from (10):

$$P^*(y|I_X = on) = P^*(y|\mathbf{r}, I_X = on) \cdot 1 \tag{11}$$

$$P^*(y) = P^*(y|\mathbf{r}) \cdot 1 \tag{12}$$

Since $Par(Y)$ blocks all paths between I_X and Y, we get $P^*(y|\mathbf{r}, I_X = on) = P^*(y|\mathbf{r})$ with the d-separation criterion, and thus, we get $P^*(y|I_X = on) = P^*(y)$ with (11) and (12). Thus, $Indep(Y, I_X = on|\mathbf{I_Z} = \mathbf{on})$ holds, which contradicts the initial assumption that $Dep(Y, I_X = on|\mathbf{I_Z} = \mathbf{on})$ holds. Therefore, (i) must be the case, i.e., there must be a path π d-connecting I_X and Y that has the form $I_X \rightarrow ... \rightarrow X \rightarrow Y$. From $\langle \mathbf{V}', E', P' \rangle$ being an i-expansion of $\langle \mathbf{V}, E, P \rangle$ it now follows that $X \rightarrow Y$ in $\langle \mathbf{V}, E \rangle$. □

Proof of theorem 3 Assume $\langle \mathbf{V}, E, P \rangle$ is an acyclic causal model and for every $X, Y \in \mathbf{V}$ (with $X \neq Y$) there is a stochastic i-expansion $\langle \mathbf{V}', E', P' \rangle$ of $\langle \mathbf{V}, E, P \rangle$ for X w.r.t. Y satisfying CMC and Min. Let X and Y be arbitrarily chosen elements of $\mathbf{V}$ such that $X \neq Y$.

(i) ⇒ (ii): Suppose $X \rightarrow Y$ in $\langle \mathbf{V}, E \rangle$. We assumed that there exists a stochastic i-expansion $\langle \mathbf{V}', E', P' \rangle$ of $\langle \mathbf{V}, E, P \rangle$ for X w.r.t. Y satisfying CMC and Min. From condition (b) of definition 13 it follows that $X \rightarrow Y$ in $\langle \mathbf{V}', E' \rangle$. Since Min is equivalent to Prod, $Dep(x, y|Par(Y)\backslash\{X\} = \mathbf{r}, I_Y = on)$ holds for some X-values x, for some Y-values y, for some of I_Y's on-values on, and for some instantiations $\mathbf{r}$ of $Par(Y)\backslash\{X\}$. Now let us assume that on is one of the I_X-values which are correlated with x conditional on $Par(Y)\backslash\{X\} = \mathbf{r}, I_Y = on$. (The existence of such an I_X-value on is guaranteed by condition (c.1) in definition 13.) Then we have $Dep(I_X = on, x|\mathbf{r}, I_Y = on) \wedge Dep(x, y|\mathbf{r}, I_Y = on)$.

We now show that $Dep(I_X = on, x|\mathbf{r}, I_Y = on) \wedge Dep(x, y|\mathbf{r}, I_Y = on)$ together with $I_X \rightarrow X \rightarrow Y$ and the d-separation criterion implies $Dep(I_X = on, y|\mathbf{r}, I_Y = on)$. We define $P^*(-)$ as $P'(-|\mathbf{r})$ and proceed as follows:

$$P^*(y|I_X = on, I_Y = on) = \sum_i P^*(y|x_i, I_X = on, I_Y = on) \cdot P^*(x_i|I_X = on, I_Y = on) \tag{13}$$

$$P^*(y|I_Y = on) = \sum_i P^*(y|x_i, I_Y = on) \cdot P^*(x_i|I_Y = on) \tag{14}$$

Equations (13) and (14) are probabilistically valid. From $I_X \rightarrow X \rightarrow Y$ and (13) we get with the d-separation criterion:

$$P^*(y|I_X = on, I_Y = on) = \sum_i P^*(y|x_i, I_Y = on) \cdot P^*(x_i|I_X = on, I_Y = on) \tag{15}$$

Since I_Y is exogenous and a causal parent only of Y, X and I_Y are d-separated by I_X, and thus, we get (16) from (15) with the d-separation criterion. Since I_Y and X are d-separated (by the empty set), we get (17) from (14) with the d-separation criterion:

$$P^*(y|I_X = on, I_Y = on) = \sum_i P^*(y|x_i, I_Y = on) \cdot P^*(x_i|I_X = on) \tag{16}$$

$$P^*(y|I_Y = on) = \sum_i P^*(y|x_i, I_Y = on) \cdot P^*(x_i) \tag{17}$$

Now either (A) $P^*(y|I_X = on, I_Y = on) \neq P^*(y|I_Y = on)$, or (B) $P^*(y|I_X = on, I_Y = on) = P^*(y|I_Y = on)$. If (A) is the case, then $Dep(Y, I_X = on|Par(Y)\backslash\{X\}, I_Y = on)$.

If (B) is the case, then $P^*(y|I_X = on, I_Y = on)$ can only equal $P^*(y|I_Y = on)$ due to a fine-tuning of $P^*(x_i|I_Y = on)$ and $P^*(x_i)$ in equations (16) and (17), respectively. We already know that X's value x and

$I_X = on$ are dependent conditional on $Par(Y)\backslash\{X\} = \mathbf{r}, I_Y = on$, meaning that $P^*(x|I_X = on, I_Y = on) \neq P^*(x|I_Y = on)$ holds. Since X and I_Y are d-separated by I_X, $P^*(x|I_X = on, I_Y = on) = P^*(x|I_X = on)$ holds. Since X and I_Y are d-separeted (by the empty set), $P^*(x|I_Y = on) = P^*(x)$ holds. It follows that $P^*(x|I_X = on) \neq P^*(x)$ holds. So (i) $P^*(x|I_X = on) > 0$ or (ii) $P^*(x) > 0$. Thanks to condition (c.2) in definition 13, every one of the conditional probabilities $P^*(y|x_i, I_Y = on)$ can be changed independently by replacing "on" in "$P^*(y|x_i, I_Y = on)$" by some I_Y-value "on^*" (with $on^* \neq on$) such that $P^*(y|x_i, I_Y = on^*) > 0$. Thus, in both cases ((i) and (ii)) it holds that $P^*(y|x, I_Y = on^*) \cdot P^*(x|I_X = on^*) \neq P^*(y|x, I_Y = on^*) \cdot P^*(x)$, while $P^*(y|x_i, I_Y = on^*) \cdot P^*(x_i|I_X = on^*) = P^*(y|x_i, I_Y = on^*) \cdot P^*(x_i)$ holds for all $x_i \neq x$. It follows that $P^*(y|I_X = on, I_Y = on^*) \neq P^*(y|I_Y = on^*)$.

(ii) ⇒ (i): Suppose $\langle \mathbf{V}', E', P' \rangle$ is one of the above assumed stochastic i-expansions for X w.r.t. Y and that $Dep(Y, I_X = on|Par(Y)\backslash\{X\}, I_Y = on)$ holds in this stochastic i-expansion. The d-separation criterion and $Dep(Y, I_X = on|Par(Y)\backslash\{X\}, I_Y = on)$ imply that I_X and Y are d-connected given $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ by a causal path $\pi : I_X - ... - Y$. π cannot have the form $I_X \leftarrow ... - Y$. This is excluded by condition (a) in **($\mathbf{IV}_S$)**. Thus, π must have the form $I_X \rightarrow ... - Y$. Now either (A) π goes through X, or (B) π does not go through X.

Suppose (B) is the case. Then, because of condition (c) in **($\mathbf{IV}_S$)**, π cannot be a directed path $I_X \rightarrow ... \rightarrow Y$. Thus, π must either (i) have the form $I_X \rightarrow ... - C \rightarrow Y$ (with a collider on π), or it (ii) must have the form $I_X \rightarrow ... - C \leftarrow Y$. If (i) is the case, then C must be in $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ (since C cannot be X). Hence, π would be blocked by $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ and, thus, would not d-connect I_X and Y given $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$. Thus, (ii) must be the case. If (ii) is the case, then there has to be a collider C^* on π that either is C or that is an effect of C, and thus, also an effect of Y. But then I_X and Y can only be d-connected given $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ over π if C^* is in $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ or has an effect in $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$. But this would mean that Y is a cause of Y, what is excluded by the initial assumption of acyclicity. Thus, (A) has to be the case.

If (A) is the case, then π must have the form $I_X \rightarrow ... - X - ... - Y$. If π would have the form $I_X \rightarrow ... - X - ... - C \leftarrow Y$ (where C and X are possibly identical), then there is at least one collider C^* lying on π that is an effect of Y. For I_X and Y to be d-connected given $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ over path π, $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ must activate π, meaning that C^* has to be in $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ or has to have an effect in $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$. But then we would end up with a causal cycle $Y \rightarrow ... \rightarrow Y$, which would contradict the assumption of acyclicity. Hence, π must have the form $I_X \rightarrow ... - X - ... - C \rightarrow Y$ (where C and X are possibly identical). Now either (i) $C = X$ or (ii) $C \neq X$. If (ii) is the case, then $C \in (Par(Y)\backslash\{X\}) \cup \{I_Y\}$, and thus, $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ blocks π. But then I_X and Y cannot be d-connected given $(Par(Y)\backslash\{X\}) \cup \{I_Y\}$ over path π. Hence, (i) must be the case. Then π has the form $I_X \rightarrow ... - X \rightarrow Y$ and from $\langle \mathbf{V}', E', P' \rangle$ being a stochastic i-expansion of $\langle \mathbf{V}, E, P \rangle$ it follows that $X \rightarrow Y$ in $\langle \mathbf{V}, E \rangle$. □