Causal exclusion and causal Bayes nets[*]

Alexander Gebharter

Abstract: In this paper I reconstruct and evaluate the validity of two versions of causal exclusion arguments within the theory of causal Bayes nets. I argue that supervenience relations formally behave like causal relations. If this is correct, then it turns out that both versions of the exclusion argument are valid when assuming the causal Markov condition and the causal minimality condition. I also investigate some consequences for the recent discussion of causal exclusion arguments in the light of an interventionist theory of causation such as Woodward's (2003) and discuss a possible objection to my causal Bayes net reconstruction.

1 Introduction

Causal exclusion arguments, most famously advanced by Kim (1989, 2000, 2003, 2005), can be used as arguments for epiphenomenalism or as arguments against non-reductive physicalism. Epiphenomenalism is the view that "mental events

[*] This is the accepted version of the following article: Gebharter, A. (2017). Causal exclusion and causal Bayes nets. *Philosophy and Phenomenological Research, 95*(2), 353–375. doi:10.1111/phpr.12247, which is published in final form at: http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1933-1592.

are caused by physical events in the brain, but have no effects upon any physical events" (Robinson, 2015). Non-reductive physicalism, on the other hand, basically consists of three assumptions: Mental properties supervene on physical properties, mental properties cannot be reduced to physical properties, and mental properties are causally efficacious (cf. Kim, 2005, p. 33).

In a nutshell, exclusion arguments assume non-reductive physicalism and conclude from several premises that mental properties supervening on physical properties cannot cause physical or other mental properties. The notion of causation used in these arguments is, however, typically somewhat vague and not specified in detail. Because of this, the validity of these arguments may depend on the specific theory of causation endorsed (cf. Hitchcock, 2012). Throughout the paper I treat theories of causation as tools for providing information about the world's true causal structure and about the causal efficacy of properties on other properties. So a theory of causation may be better w.r.t. providing such information than another theory. If two such theories lead to different results about the validity of exclusion arguments, then the one providing more information relevant for exclusion arguments should be favored when evaluating the validity of such arguments.

In this paper I reconstruct two versions of exclusion arguments and evaluate their validity within a particular theory of causation, *viz.* the theory of causal Bayes nets. The theory of causal Bayes nets (CBNs) evolved from the Bayes net formalism (Neapolitan, 1990; Pearl, 1988). It was elaborated in detail by researchers such as Pearl (2000) and Spirtes, Glymour, and Scheines (2000). The theory connects causal structures to probability distributions and provides powerful methods for causal discovery, prediction, and testing of causal hypotheses. Furthermore, its core axioms can be justified by an inference to the best explanation (see Schurz, 2008 for a general approach) of certain statistical phenomena, and several versions of the theory can be proven to have empirical content, by whose means not only the theory's models, but also the theory as a whole becomes empirically testable (Schurz & Gebharter, 2015). So the theory of CBNs probably gives us the best empirical grasp on causation we have so far. Hence, it allows for an empirically informed treatment of causation in causal exclusion arguments, and thus, also for an empirically informed evaluation of the validity of such arguments.

Another strong motivation for this endeavor is that causal exclusion arguments have recently been intensively discussed (cf., e.g., Baumgartner, 2009, 2010; Eronen, 2012; Raatikainen, 2010; Shapiro, 2010; Shapiro & Sober, 2007; Woodward, 2008, 2014) within an interventionist framework of causation à la Woodward (2003), and that interventionist accounts do have a natural counterpart within the theory of CBNs (cf., e.g., Gebharter & Schurz, 2014 or Zhang & Spirtes, 2011). So the hope is that we can draw as of yet unconsidered conclusions for the interventionist debate surrounding causal exclusion arguments from a reconstruction on the basis of the theory of CBNs. This seems especially promising since one of the main problems interventionists have when testing causal efficacy of properties standing in supervenience relationships to other properties is that these properties cannot be simultaneously manipulated by interventions (for details, see section 4). So the interventionist account seems to have some kind of a blind spot when it comes to testing causal efficacy of such properties. The theory of CBNs, on the other hand, provides a neat and simple test for causal efficacy not requiring fixability by means of interventions.

The paper is structured as follows: In section 2 I briefly introduce two variants of the causal exclusion argument. In section 3, which is the main section of the paper, I reconstruct these two variants within the theory of CBNs and evaluate their validity. This requires an answer to the question of how supervenience relationships should be represented in CBNs and a test for evaluating whether the instantiation of a property X at least sometimes contributes something to the occurrence of another property Y. I will argue that supervenience relationships can be treated similar to a CBN's causal arrows. This assumption will be crucial for my argumentation in subsequent sections. A method for testing a property's causal efficacy is already implemented in the productivity condition, which can be proven to be equivalent to one of the theory of CBN's core axioms, *viz.* the causal minimality condition (cf. Spirtes et al., 2000, p. 31). I conclude section 3 by demonstrating that mental properties supervening on physical properties cannot be causally efficacious if causal as well as supervenience relations are assumed to obey the core axioms of the theory of causal nets. In section 4 I investigate the consequences of these findings for the interventionist debate on the causal exclusion argument. In section 5 I defend my suggestion to treat supervenience relationships similar to causal arrows against an objection raised by Woodward (2014). I conclude in section 6.

2 The causal exclusion argument

Causal exclusion arguments (cf. Kim, 1989, 2000, 2003, 2005) typically come in two variants (cf. Harbecke, 2013): (i) arguments against the causal efficacy of mental properties on physical properties, and (ii) arguments against the causal efficacy of mental properties on other mental properties. In this paper we will have a look at both variants.

The diagram in Figure 1 (which is adapted from Kim, 2005) can be used to illustrate both versions of the causal exclusion argument. P_1 and P_2 stand for physical properties, while M_1 and M_2 stand for mental properties. P_1, P_2, M_1, and M_2 are assumed to be pairwise non-identical. Furthermore, we also assume that there is no spatio-temporal overlap of P_1 and P_2. Double-tailed arrows

$$\begin{tikzcd}
M_1 \arrow[r, "?"] \arrow[dr, "?"] & M_2 \\
P_1 \arrow[u, Rightarrow] \arrow[r] & P_2 \arrow[u, Rightarrow]
\end{tikzcd}$$

Figure 1: Diagram for illustrating the two versions of the causal exclusion argument. Single-tailed arrows stand for direct causal relations, while double-tailed arrows indicate supervenience relationships.

($\Longrightarrow$) represent relationships of supervenience. So M_1 supervenes on P_1, and M_2 supervenes on P_2, meaning that every change in M_1 and M_2 is necessarily associated with a change in P_1 and P_2, respectively (cf. McLaughlin & Bennett, 2011). In addition, we assume that P_1 and P_2 fully determine M_1 and M_2, respectively. So the occurrence of P_1 and P_2 suffices for the instantiation of M_1 and M_2, respectively. Alternatively we can say that P_1 constitutes M_1 and that P_2 constitutes M_2.[1]

Single-tailed arrows ($\longrightarrow$) represent direct causal relationships. So P_1 is a direct cause of P_2. Because we assume the completeness of the physical domain, i.e., that every physical property has a sufficient physical cause, also P_2 has a sufficient physical cause.[2] We assume P_1 to be that cause, and hence, P_1's occurrence determines P_2's occurrence. Now the question is whether we can

[1] The properties I called supervenience and constitution here are typically combined by assuming strong supervenience (cf. Kim, 2003, p. 151). However, I prefer to separate them in this paper.

[2] According to the completeness of the physical, also P_1 will have a sufficient physical cause. Since P_1's physical causes, however, will not be relevant for the argument, we do not represent them in the diagram.

draw single-tailed arrows from M_1 to P_2 and from M_1 to M_2, i.e., whether M_1 can cause P_2 or M_2. This question is represented by the question marks over the single-tailed arrows $M_1 \longrightarrow P_2$ and $M_1 \longrightarrow M_2$ in the diagram.

Version (i) of the causal exclusion argument roughly goes as follows: P_1, P_2, M_1, and M_2 are instantiated. Now let us ask why P_2 is instantiated. Because of the causal completeness of the physical, P_1's instantiation suffices for P_2's instantiation. So P_2 is instantiated because P_1 is. Since P_1's occurrence necessitates P_2's occurrence, there is nothing the instantiation of M_1 could contribute to P_2's occurrence. Hence, M_1 has no causal influence on P_2.

Version (ii) of the argument roughly goes as follows: P_1, P_2, M_1, and M_2 are instantiated. Now the crucial question is why M_2 is instantiated. M_2 is constituted, and thus, fully determined by its physical supervenience base P_2. So P_2's occurrence suffices for M_2's occurrence. Hence, M_2 is instantiated because P_2 is. Since P_2's occurrence necessitates M_2's occurrence, there is nothing left the instantiation of M_1 could contribute to M_2's occurrence. Thus, M_1 cannot cause M_2.[3]

One assumption both versions of the exclusion argument require is that a cause's instantiation contributes at least sometimes something to the occurrence of its direct effects. This assumption is highly plausible in the light of Occam's razor, which states that one should assume theoretical entities (e.g., direct causal relations) only when they are required to explain otherwise unexplainable empirical facts. It will play a major role in the reconstruction of the

[3] I am indebted to Wlodek Rabinowicz for pointing out to me that one could conclude that P_1 cannot be a cause of M_2 by a similar argumentation, which even epiphenomenalists might find counterintuitive. However, the epiphenomenalist could solve this problem by interpreting constitution as a causal relation: She could then conclude that P_1 cannot be a direct cause of M_2, but that P_1 can be an indirect cause of M_2. P_1 first directly causes P_2, which then directly causes M_2. Note that M_1—contrary to P_1—cannot even be an indirect cause of M_2.

causal exclusion argument in terms of causal Bayes nets.

3 Causal exclusion and causal Bayes nets

In this section I reconstruct both versions of the causal exclusion argument and evaluate their validity on the basis of the empirically well-informed theory of causal Bayes nets. I start with introducing important notions and the core axioms of the theory of CBNs. A causal model is a triple $\langle \mathbf{V}, \mathbf{E}, P \rangle$ in which $\mathbf{V}$ is a set of variables, $\langle \mathbf{V}, \mathbf{E} \rangle$ is a directed acyclic graph (DAG) over $\mathbf{V}$, and P is a probability distribution over $\mathbf{V}$. The DAG $\langle \mathbf{V}, \mathbf{E} \rangle$ represents the modeled system's causal structure, where $X_i \longrightarrow X_j$ means that X_i is a direct cause of X_j (w.r.t. $\mathbf{V}$). The set of all direct causes of a variable X_i in a causal model is called the set of X_i's parents $\mathbf{Par}(X_i)$. The union of the set of all effects of a variable X_i (i.e., the set of all X_j with $X_i \longrightarrow ... \longrightarrow X_j$) in a causal model and $\{X_i\}$ is called the set of X_i's descendants $\mathbf{Des}(X_i)$.

A causal model's probability distribution P represents the causal strengths of the causal influences propagated along the causal arrows. We define probabilistic dependence of a variable X on another variable Y conditional on a variable (or a set of variables) Z—$Dep(X, Y|Z)$ for short— as $P(x|y,z) \neq P(x|z) \wedge P(y,z) > 0$ for some X-, Y-, and Z-values x, y, and z, respectively. X's probabilistic independence from Y conditional on Z—$Indep(X, Y|Z)$ for short—is defined as the negation of $Dep(X, Y|Z)$, i.e., as $P(x|y,z) = P(x|z) \vee P(y,z) = 0$ for all X-, Y-, and Z-values x, y, and z, respectively.

The first axiom of the theory of CBNs is the causal Markov condition (CMC). A causal model $\langle \mathbf{V}, \mathbf{E}, P \rangle$ satisfies CMC if and only if every variable X_i in $\mathbf{V}$ is probabilistically independent of its non-descendants conditional on its direct causes (Spirtes et al., 2000, p. 29). CBNs are causal models that satisfy CMC.

The DAG of a CBN determines the following Markov factorization:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|\mathbf{Par}(X_i)) \tag{1}$$

Another important axiom is the causal minimality condition (Min). A CBN $\langle \mathbf{V}, \mathbf{E}, P \rangle$ satisfies (Min) if and only if there is no CBN $\langle \mathbf{V}, \mathbf{E}', P \rangle$ with $\mathbf{E}' \subset \mathbf{E}$ (cf. Spirtes et al., 2000, p. 31). In other words: If deleting some causal arrow of the CBN's graph would lead to a causal model that violates CMC, then the CBN is minimal, i.e., every arrow is required to prevent some (conditional) independence relation. This means that every arrow is responsible for some probabilistic dependence between the variables it connects. The idea that every causal arrow leaves some probabilistic footprints can also be expressed by the following productivity condition (Prod), which can be proven to be equivalent to Min for CBNs (Gebharter & Schurz, 2014, Theorem 1):

> **Prod** A causal model $\langle \mathbf{V}, \mathbf{E}, P \rangle$ satisfies Prod if and only if $Dep(X_j, X_i|\mathbf{Par}(X_j)\backslash\{X_i\})$ holds for all $X_i, X_j \in \mathbf{V}$ with $X_i \longrightarrow X_j$.

Recall from section 2 that causal exclusion arguments presuppose that a cause contributes at least sometimes something to the occurrence of its direct effects. This somewhat vaguely formulated requirement can be stated more precisely by means of a condition implemented in Prod above: A causal arrow $X_i \longrightarrow X_j$ in a CBN is productive (or, equivalently, X_i contributes at least sometimes something to X_j) if and only if $Dep(X_j, X_i|\mathbf{Par}(X_j)\backslash\{X_i\})$ holds, i.e., if there are some X_i- and X_j-values x_i and x_j, respectively, such that X_i's taking value x_i makes a probabilistic difference for X_j's taking value x_j w.r.t. some fixed context $\mathbf{Par}(X_j)\backslash\{X_i\} = \mathbf{r}$. We can use this insight to test specific causal arrows for whether they are productive.

Now we also see how Prod can be used to nicely reflect the assumption (made in both versions of the causal exclusion argument) that properties can cause other properties only if they can contribute something to the occurrence of the latter. In causal Bayes nets terminology this assumption means to only allow for minimal CBNs, i.e., for CBNs that satisfy Prod. Or in other words: Only productive arrows can represent real causal relations.

Let us try to reconstruct both versions of the causal exclusion argument on the basis of the theory of CBNs next. To this end, let our CBN's variable set $\mathbf{V}$ be identical to $\{P_1, P_2, M_1, M_2\}$. How does our CBN's graph have to look like? We have to draw an arrow from P_1 to P_2, since P_1 is assumed to directly cause P_2. We also have to draw arrows from M_1 to P_2 and from M_1 to M_2. These latter two arrows are the ones we want to test for productivity. But how should we represent the double-tailed arrows (indicating supervenience relationships) in our CBN? As Woodward (2014, sec. 1) remarks, there has been little discussion in the causal modeling literature on how to represent supervenience relationships (or other non-causal relationships). The answer to this question is not trivial. I will suggest an answer that requires that we take a closer look at how arrows work in causal Bayes nets. In particular, I will suggest to treat supervenience relationships similar to causal arrows in CBNs. In my argumentation I will not make use of intuitions about how interventions should work together with supervenience relationships. The reason for this is that there is still no consensus about that. It is still unclear whether directly intervening on mental properties is possible, or whether mental properties can only be manipulated by intervening on their physical supervenience bases, whether interventions on mental properties have to be common causes of these mental properties and their supervenience bases, etc. One of the goals of this paper is to implement supervenience dependencies in causal models to answer questions like these. Hence, I cannot let intuitions about how interventions should work together with supervenience relations enter my argumentation for how to represent them in CBNs.

A nice feature of CBNs is that, due to Equation 1, the conditional probabilities $P(X_i|\mathbf{Par}(X_i))$ corresponding to the CBN's arrows—these conditional probabilities are also called the CBN's parameters—are stable in the sense that they do not vary when one changes the prior distribution of some non-descendants of X_i. Let me illustrate this by means of the following simple example: Assume a CBN with DAG $X \longrightarrow Y$, where X and Y are binary variables. According to Equation 1, the CBN's probability distribution factors as $P(X,Y) = P(Y|X) \cdot P(X)$. Let us assume $P(Y|X)$ and $P(X)$ are specified as follows:

$$\begin{array}{ll} P(x_1) = 0.25 & P(y_1|x_1) = 0.75 \\ P(x_0) = 0.75 & P(y_0|x_1) = 0.25 \\ & P(y_1|x_0) = 0.5 \\ & P(y_0|x_0) = 0.5 \end{array}$$

Changing X's prior distribution will not have an influence on the model's parameters $P(Y|X)$. It will, however, typically change conditional probabilities which are non-parameters. $P(x_1|y_1)$, for example, is such a non-parameter conditional probability. It can be computed as follows:

$$P(x_1|y_1) = \frac{P(y_1|x_1) \cdot P(x_1)}{\sum_{x_i} P(y_1|x_i) \cdot P(x_i)} = \frac{0.75 \cdot 0.25}{0.5625} = 0.\dot{3} \qquad (2)$$

Now, if we change $P(x_1)$ from 0.25 to 0.5, for example, we get a different conditional probability $P(x_1|y_1)$:

$$P(x_1|y_1) = \frac{P(y_1|x_1) \cdot P(x_1)}{\sum_{x_i} P(y_1|x_i) \cdot P(x_i)} = \frac{0.75 \cdot 0.5}{0.625} = 0.6 \qquad (3)$$

The stability of a CBN's parameters explained above corresponds to the intuition that each subsystem of a CBN consisting of a variable X_i and its direct causes $\mathbf{Par}(X_i)$ represents an autonomous causal mechanism (cf. Pearl, 2000, p. 22). Supervenience relationships, as assumed in the causal exclusion argument, seem to also possess this stability property. Recall that every mental property M_i is constituted by some physical property (or properties) P_i. This means that for every P_i-value p_i there has to be exactly one M_i-value m_i such that $P(m_i|p_i) = 1$ holds, where the conditional probabilities $P(m_i|p_i) = 1$ cannot be changed by changing the prior distribution of some non-descendants of M_i. Moreover, the conditional probabilities $P(m_i|p_i) = 1$ cannot even be changed by modifying the prior distribution of non-descendants of M_i in any possible expansion of our CBN. Otherwise P_i would not constitute M_i. Constitution is a metaphysical notion. If we would find that P_i does not determine M_i anymore when modifying the prior distribution of some non-descendants of M_i in an expansion of the CBN, we would conclude that we falsely took M_i as constituted by P_i. We would conclude that we found some kind of dependence of M_i on P_i that just looks like constitution in certain circumstances.

There are (at least) two further reasons for treating supervenience relationships $P_i \Longrightarrow M_i$ like causal arrows in a CBN. The first one is that this representation fits the idea of multiple realizability nicely, which typically goes hand in hand with the assumption that mental properties supervene on physical properties. For our causal diagram supervenience means that every M_i-value change has to lead to some probability change for some P_i-value. But the conditional probabilities $P(P_i|M_i)$ do not have to equal 1 or 0. They can vary when M_i's value is fixed. This directly corresponds to the multiple realizability intuition. Sometimes P_i can be changed while M_i is fixed, or in other words: There may be several instantiations of P_i that all constitute a certain instantiation of M_i.

The last point speaking for treating a supervenience relationship $P_i \Longrightarrow M_i$

similar to a causal relationship in a CBN is that micro properties are oftentimes understood as causes of macro properties. Friends of non-reductive physicalism could, for example, describe the temperature of a gas in a tank as the effect of the behavior of the gas particles in the tank, etc.

Summarizing, we found some good reasons to treat double-tailed arrows ($\Longrightarrow$) standing for supervenience relationships like a CBN's causal arrows ($\longrightarrow$).[4] Thus, our CBN also has to feature a double-tailed arrow from P_1 to M_1 and from P_2 to M_2. So the DAG of our CBN ultimately turns out to be the graph depicted in Figure 1, where the double-tailed arrows technically work exactly like single-tailed arrows. Note that I do not want to claim that supervenience is a special kind of causation here. I rather prefer to stay neutral on that question. My claim is that supervenience relationships, since they have the same formal properties as causal relations, can be modeled and formally represented in CBNs similar to causal relations. Woodward (2014) raised an objection from the perspective of an interventionist theory of causation to treating supervenience relationships like causal arrows. I illustrate his objection and defend my suggestion of how to model supervenience in CBNs in section 5.

[4] I would like to acknowledge that there may still be good reasons for not treating supervenience relationships like causal relationships in CBNs. So the adequacy of my suggestion is still debatable. Also note that within an interventionist framework (such as Woodward's 2003) the consequences of the exclusion argument follow straightforwardly if one accepts that supervenience relations behave like ordinary causal relations in CBNs. Woodward (2014) is aware of this fact and no very elaborate further analysis is required to show this. (See section 4 for details.) I am indebted to an anonymous referee for this remark. However, within the CBN framework one would need to add something like Spirtes and Zhang's (2011, p. 338) intervention principle to get these consequences similarly and in a straightforward interventionist fashion. The alternative way to do it consists in using the productivity test I introduced earlier. Using this productivity test has the advantage over introducing an additional intervention principle that it does not make the theory stronger than it has to be.

We now know how our CBN's DAG has to look like. But how should we specify its probability distribution P? From the considerations above we also already know that: Assuming the completeness of the physical domain means that there is a sufficient physical cause for every physical property. Physical properties are represented by the variables P_1 and P_2 in our CBN. We assume P_1 to represent P_2's sufficient physical cause, meaning that every P_1-value p_1 is sufficient for P_2 taking a certain value p_2. This probabilistic constraint that comes with assuming the completeness of the physical is labeled "physical completeness" below.

Our next constraint comes with assuming that every mental property supervenes on some physical property. Mental properties are represented by the variables M_1 and M_2 in our CBN. M_1 is assumed to supervene on P_1, and M_2 is assumed to supervene on P_2. That M_i supervenes on P_i means in probabilistic terms that every value change of M_i has to lead to a probability change for some P_i-values. This constraint is labeled "supervenience" below.

The last constraint comes with the assumption that mental properties are constituted by physical properties. We assume that P_1 constitutes M_1, and that P_2 constitutes M_2. Since the constituting property determines the constituted property, we have to assume for our CBN's probability distribution that every value p_1 of P_1 determines M_1 to take a certain value m_1, and that every value p_2 of P_2 determines M_2 to take a certain value m_2. This constraint is labeled "constitution" below.

Summarizing, our CBN's probability distribution P has to satisfy the following probabilistic requirements, where $i \in \{1, 2\}$:

Physical completeness $\forall p_1 \exists p_2 : P(p_2|p_1) = 1$

Supervenience $\forall m_i \forall m_i' \exists p_i : m_i \neq m_i' \rightarrow P(p_i|m_i) \neq P(p_i|m_i')$

Constitution $\forall p_i \exists m_i : P(m_i|p_i) = 1$

Now version (i) of the causal exclusion argument concludes that M_1 cannot cause P_2, since P_2 is fully determined by P_1 and M_1 has nothing to contribute to whether P_2 occurs. We get the same result from our CBN. We can test the causal productiveness of the arrow $M_1 \longrightarrow P_2$ by checking whether $Dep(P_2, M_1|\mathbf{Par}(P_2)\backslash\{M_1\})$ holds. Let p_1 be an arbitrarily chosen P_1-value. Due to the completeness of the physical for every p_1 there is exactly one p_2 such that $P(p_2|p_1) = 1$ holds, while $P(p_2'|p_1) = 0$ holds for all $p_2' \neq p_2$. Now for every m_1 there are two possible cases.

Case 1: m_1 and p_1 are compatible, i.e., $P(m_1, p_1) > 0$ holds. In that case conditionalizing on m_1 will not change the conditional probabilities of p_2 or p_2' given p_1, i.e., also $P(p_2|m_1, p_1) = 1$ and $P(p_2'|m_1, p_1) = 0$ will hold, meaning that no P_2-value depends on m_1 conditional on p_1.

Case 2: m_1 and p_1 are incompatible, i.e., $P(m_1, p_1) = 0$ holds. It then follows from the definition of probabilistic independence introduced earlier that no P_2-value depends on m_1 conditional on p_1.

It follows that conditionalizing on p_1 will render P_2 independent from M_1. Since p_1 was arbitrarily chosen, we can generalize this result: Conditionalizing on any P_1-value p_1 will render P_2 independent from M_1, i.e., P_2 and M_1 are independent conditional on $\mathbf{Par}(P_2)\backslash\{M_1\} = P_1$, meaning that the arrow $M_1 \longrightarrow P_2$ is unproductive.

Since P_1 is assumed to physically (or nomologically) determine P_2, this result can be generalized for every possible expansion of our CBN, meaning that M_1 cannot cause P_2 in any circumstances.

Version (ii) of the exclusion argument says that M_1 cannot cause M_2, since M_2 is fully determined by P_2 and, hence, M_1 cannot contribute anything to whether M_2 occurs. This claim is also provable within our CBN. Again, we can test the causal productiveness of the arrow $M_1 \longrightarrow M_2$ by checking whether

$Dep(M_2, M_1|\mathbf{Par}(M_2)\backslash\{M_1\})$ holds. Let p_2 be an arbitrarily chosen P_2-value. Due to the fact that P_2 constitutes M_2, there is exactly one M_2-value m_2 for every P_2-value p_2 such that $P(m_2|p_2) = 1$ holds, while $P(m_2'|p_2) = 0$ holds for all $m_2' \neq m_1$. Now for every M_1-value m_1 there are two possible cases.

Case 1: m_1 and p_2 are compatible, meaning that $P(m_1, p_2) > 0$ holds. Then $P(m_2|m_1, p_2) = 1$ and $P(m_2'|m_1, p_2) = 0$ will hold. Hence, no M_2-value depends on m_1 conditional on p_2.

Case 2: m_1 and p_2 are incompatible, meaning that $P(m_1, p_2) = 0$ holds. From this it follows, again by the definition of probabilistic independence, that no M_2-value depends on m_1 conditional on p_2.

Therefore, conditionalizing on p_2 renders M_2 probabilistically independent from M_1. Recall that p_2 was arbitrarily chosen. Hence, we can generalize our result: Conditionalizing on any P_2-value will render M_2 probabilistically independent from M_1, meaning that M_2 and M_1 are independent conditional on $\mathbf{Par}(M_2)\backslash\{M_1\} = P_2$. Thus, the arrow $M_1 \longrightarrow M_2$ is unproductive.

Since P_2 is assumed to constitute M_2, this result can be generalized for every possible expansion of our CBN. This means, again, that M_1 cannot cause M_2 in any circumstances.

The argumentation for the unproductiveness of the causal arrow $M_1 \longrightarrow M_2$ basically follows the same pattern as the one for the unproductiveness of the causal arrow $M_1 \longrightarrow P_2$: In both cases we have a substructure $X \longrightarrow Z \longleftarrow Y$ such that for every Y-value y there is exactly one Z-value z such that $P(z|y) = 1$ holds, while $P(z'|y) = 0$ holds for all $z' \neq z$. For every Y-value y we distinguish the X-values x which are compatible with y from the ones which are not. Conditionalizing on compatible X-values in addition to y—this is case 1 above—will not change the probabilities of z and z', meaning that also $P(z|x, y) = 1$ and $P(z'|x, y) = 0$ will hold. So conditionalizing on a Y-value y

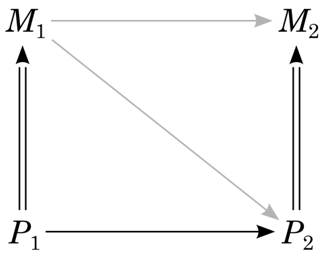

Figure 2: Our productivity test reveals that the grey arrows $M_1 \longrightarrow P_2$ and $M_1 \longrightarrow M_2$ cannot propagate probabilistic dependence between the variables at their heads and tails.

renders Z independent of those X-values x compatible with y. But Z will also be rendered independent of those X-values x which are incompatible with y—this is case 2 above. The latter follows directly from the definition of probabilistic independence and $P(x, y) = 0$.

Figure 2 summarizes our findings in this section: The grey arrows $M_1 \longrightarrow P_2$ and $M_1 \longrightarrow M_2$ turn out to be unproductive, meaning that they cannot propagate probability between M_1 and P_2 or M_2. This result holds for every possible expansion of our CBN. If one accepts the plausible assumption that only those properties X are causes of a property Y, which have at least sometimes some probabilistic influence on Y (i.e., if one assumes Prod), then we have to delete the two grey arrows.[5]

Our result may be interpreted as empirically informed support for epiphenomenalism or as evidence against non-reductive physicalism: If causation is

[5] Note that it is well-known that the causal minimality condition (and, hence, also Prod) may be violated in CBNs whose probability distributions feature deterministic dependencies (cf. Zhang & Spirtes, 2011). The unproductiveness of the causal arrows $M_1 \longrightarrow P_2$ and $M_1 \longrightarrow M_2$ in our causal exclusion CBN is basically such a violation of the causal minimality condition due to deterministic dependencies.

characterized by means of the causal Markov condition and the causal minimality condition, we assume that mental properties are non-identical to their physical supervenience bases, and that every physical property has a sufficient physical cause, then mental properties cannot act as causes for physical properties or as causes for other mental properties—they possess no causal power. This result may, however, also be interpreted as evidence against non-reductive physicalism: By dropping the assumption that mental properties are non-identical to their supervenience bases we can restore mental properties' causal efficacy. In that case we would not represent mental and physical properties by different variables in our causal model, and hence, our causal graph would become "flat": M_1 would be identical to P_1 and M_2 would be identical to P_2 and we can, of course, have a productive arrow from $M_1 = P_1$ to $M_2 = P_2$.

4 Consequences for the causal exclusion debate within the interventionist framework

Shapiro and Sober (2007) and others (e.g., Raatikainen, 2010; Shapiro, 2010; Woodward, 2008) have recently argued that causal exclusion arguments are not valid in the light of an interventionist theory of causation such as Woodward's (2003). This started a still ongoing debate within the interventionist framework about whether mental properties can (in principle) be efficacious causes of physical and other mental properties and about whether the causal efficacy of such mental properties can be accounted for on empirical grounds if the answer to the former question is a positive one.

Within an interventionist theory of causation such as Woodward's (2003), the efficacy of single causal arrows $X \longrightarrow Y$ can only be tested by means of interventions.[6] The method used for testing whether $X \longrightarrow Y$ is causally efficacious is basically the same as the method for testing whether X is a direct cause of Y: One has to fix all elements of one's set of variables $\mathbf{V}$ of interest different from X and Y by interventions and check whether Y would change when manipulating X (cf. Woodward, 2003, p. 59). Baumgartner (2010, 2009) convincingly showed that one gets problems with this test for causal efficacy in the presence of supervenience relationships since P_1's value cannot be fixed by an intervention while M_1's value is changed by another intervention, simply because M_1 supervenes on P_1. Hence, Woodward's (2003) interventionist account would not only yield that M_1 cannot be causally efficacious, but also that M_1 cannot even be a direct cause of P_2 or M_2. Baumgartner discusses several possibilities to modify Woodward's (2003) interventionist theory to solve this problem, but concludes that none of these modifications would provide empirical evidence for mental properties' causal efficacy, which is what the non-reductive physicalist wants. Later on Woodward (2014) made explicit that the version of his interventionist theory of causation he presented in (Woodward, 2003) was not intended to be applied to sets containing variables standing in non-causal relationships (such as supervenience relationships). For such variable sets Woodward (2014) proposes a modified interventionist theory that does not require to fix variables standing in non-causal dependencies to X or Y by interventions when testing for whether X is a direct cause of Y. He argues that this move in principle allows mental properties to be causally efficacious w.r.t. physical properties or other mental properties.

[6] One may argue that Woodward's (2003) interventionist theory does not exclude tests for whether X is a direct cause of Y based on observational data. But this would require additional principles, such as the causal Markov condition etc., not included in Woodward's original interventionist theory. I take it that adding such principles would commit one to a CBN framework.

Baumgartner (2013) highlights the following shortcoming of Woodward's (2014) modified interventionist theory: One consequence of this theory is that any intervention $I_{M_1} = i_{M_1}$ on the mental property M_1 has to cause both M_1 and its supervenience base P_1 over two different causal paths ($M_1 \longleftarrow I_{M_1} \longrightarrow P_1$).[7] So we get M_1 as a direct cause of P_2 if some intervention $I_{M_1} = i_{M_1}$ leads to a change in P_2. But can we also show that $M_1 \longrightarrow P_2$ is a productive causal relation—which is what the non-reductive physicalist wants—within Woodward's modified account? The only possibility to test M_1's direct causal efficacy on P_2 we have is to block all directed paths from I_{M_1} to P_2 different from $I_{M_1} \longrightarrow M_1 \longrightarrow P_2$ by interventions and check whether some intervention $I_{M_1} = i_{M_1}$ leads to a change in P_2. But we already know what happens if we carry out this test: Intervening on P_1 freezes M_1 to a certain value and no intervention $I_{M_1} = i_{M_1}$ can lead to a change in P_2 anymore. There is no other way to test whether $M_1 \longrightarrow P_2$ is productive within an interventionist framework. So it is, at least in principle, possible that the change in P_2 associated with $I_{M_1} = i_{M_1}$ is solely due to the causal path $I_{M_1} \longrightarrow P_1 \longrightarrow P_2$, while the arrow $M_1 \longrightarrow P_2$ is unproductive. Hence, though Woodward's modified interventionist theory leads to the consequence that M_1 is a direct cause of P_2, it cannot lend any support to the efficacy of M_1 on P_2. One can formulate an analogous argument for the causal arrow $M_1 \longrightarrow M_2$. So Woodward's modified interventionist theory seems to have some kind of a blind spot when it comes to determining whether arrows exiting variables whose supervenience bases are also included in the variable set of interest are causally efficacious.

However, one may argue (as Woodward, 2014 does) that the modified interventionist account still renders causal exclusion arguments invalid, since it (at

[7] Such common cause interventions are called fat-handed interventions in (Baumgartner & Gebharter, 2015).

least in principle) allows for M_1 to be causally efficacious. It is at this point where we can enter the debate. We can say more about M_1's causal efficacy than Woodward can: The productivity test carried out in section 3 yields that both arrows are unproductive. So it turns out that both versions of the exclusion argument are valid when modeled in a CBN framework, while they are invalid when applying Woodward's modified interventionist theory of causation.

Which consequences should we draw from this observation for our two versions of the exclusion argument? Recall from section 1 that both theories are treated as tools for providing information about the world's true causal structure as well as the causal efficacy of properties on other properties. Such tools may be better or not so good in providing such information. In case the set of variables $\mathbf{V}$ whose causal structure should be analyzed contains variables standing in relationships of supervenience, it turned out that the theory of CBNs provides more information about the causal efficacy of some variables. It can be used for determining the causal efficacy of variables whose supervenience bases are included in $\mathbf{V}$ by means of the productivity test introduced in section 3, while Woodward's (2014) modified interventionist theory keeps silent about the causal efficacy of such variables. But in evaluating the validity of the two versions of the exclusion argument we should use as much relevant information as possible. Hence, we should conclude that both argument versions are invalid within Woodward's modified interventionist framework only because this framework does not come with a test for causal efficacy applicable to mental properties supervening on physical properties. But if we apply the CBN framework we get the relevant information that the arrows $M_1 \longrightarrow P_2$ and $M_1 \longrightarrow M_2$ are unproductive. Thus, we should conclude that both argument versions are valid.

Now, as a final step, let us see what happens if we assume, as interventionists do, that there exists an intervention variable I_{M_1} for M_1 that is correlated with

P_2 within our CBN representation. We add this intervention variable as a direct cause of M_1 to our model. Now, since $M_1 \longrightarrow P_2$ is unproductive, we can conclude that I_{M_1} can definitely not influence P_2 over path $I_{M_1} \longrightarrow M_1 \longrightarrow P_2$. Hence, I_{M_1} must influence P_2 over another path. Thus, I_{M_1} must be a direct common cause of M_1 and of at least one of the variables P_1, P_2, or M_2. Since P_2 is fully determined by P_1 (completeness of the physical) and M_2 is fully determined by P_2 (constitution), the argumentation pattern described in section 3 can be applied and our productivity test will lead to the result that arrows $I_{M_1} \longrightarrow P_2$ and $I_{M_1} \longrightarrow M_2$ would be unproductive. Hence, the correlation between I_{M_1} and P_2 can only be due to a causal path $I_{M_1} \longrightarrow P_1 \longrightarrow P_2$, and thus, I_{M_1} must be a (direct) common cause of M_1 and P_1 such that $I_{M_1} \longrightarrow P_1$ is productive. So up to this point, we arrive at a consequence close to Baumgartner's (2013). The main difference is that Baumgartner only showed that it is possible within the modified interventionist framework that I_{M_1} influences P_2 solely over path $I_{M_1} \longrightarrow P_1 \longrightarrow P_2$. We were also able to show that $I_{M_1} \longrightarrow P_1 \longrightarrow P_2$ is the only path over which I_{M_1} can be efficacious w.r.t. P_2.

But we can say even more: The argumentation pattern used to show the unproductiveness of the arrows $M_1 \longrightarrow P_2$ and $M_1 \longrightarrow M_2$ in section 3 can also be applied to $I_{M_1} \longrightarrow M_1$: Let p_1 be an arbitrarily chosen P_1-value. Since M_1 is constituted by P_1, for every P_1-value p_1 there is exactly one M_1-value m_1 such that $P(m_1|p_1) = 1$, while $P(m_1'|p_1) = 0$ holds for all $m_1' \neq m_1$. Now for every I_{M_1}-value i_{M_1} there are two possible cases.

Case 1: i_{M_1} and p_1 are compatible, which means that $P(i_{M_1}, p_1) > 0$. If this is the case, then also $P(m_1|i_{M_1}, p_1) = 1$ and $P(m_1'|i_{M_1}, p_1) = 0$ will hold. Therefore, no M_1-value will depend on i_{M_1} conditional on p_1.

Case 2: i_{M_1} and p_1 are incompatible, i.e., $P(i_{M_1}, p_1) = 0$ holds. Then it directly follows from the definition of probabilistic independence that no M_1-value probabilistically depends on i_{M_1} conditional on p_1.

Therefore, M_1 is independent from I_{M_1} when conditionalizing on p_1. Since p_1 was arbitrarily chosen, we can, again, generalize this result: Conditionalizing on any P_1-value will render M_1 independent from I_{M_1}, meaning that the causal arrow $I_{M_1} \longrightarrow M_1$ is unproductive.[8]

This result can be generalized for every possible expansion of our CBN simply because P_1 is assumed to constitute M_1 and constitution is a metaphysical notion. Hence, I_{M_1} cannot be a productive direct cause of M_1 in any circumstances.

Figure 3 graphically illustrates our findings, where single-tailed grey arrows, again, represent unproductive (possible) causal relations. For testing M_1's causal efficacy on P_2 and M_2 interventionists require interventions $I_{M_1} = i_{M_1}$ on M_1 which at least sometimes make a difference for M_1. But, as demonstrated, an intervention variable I_{M_1} for M_1 can only stand to P_1 in a productive direct causal relationship. It follows that if causation is characterized by CMC and

[8] One may object that it seems that the productivity test suggested will not work in case M_1 and P_1 are perfectly correlated, meaning that every M_1-value determines a certain P_1-value (with probability 1) and vice versa. In such a scenario one could ask why the arrow $I_{M_1} \longrightarrow M_1$ and not the arrow $I_{M_1} \longrightarrow P_1$ should be regarded as unproductive. If M_1 and P_1 are perfectly correlated, not only conditionalizing on P_1 will render M_1 independent of I_{M_1}, but also conditionalizing on M_1 will render P_1 independent of I_{M_1}. Here is my response: Who argues in such a way seems to have overlooked that the productivity test suggested also makes use of the system of interest's underlying structure. For testing whether $I_{M_1} \longrightarrow M_1$ is productive we have, according to our productivity test, to check whether M_1 depends on I_{M_1} conditional on its parents different from I_{M_1}, i.e., conditional on P_1. For testing whether $I_{M_1} \longrightarrow P_1$ is productive, on the other hand, we have to check whether P_1 depends on I_{M_1} unconditionally (since P_1 does not have any parents different from I_{M_1}). We find $\mathit{Indep}(M_1, I_{M_1}|P_1)$ and $\mathit{Dep}(P_1, I_{M_1})$ and conclude that $I_{M_1} \longrightarrow M_1$ is unproductive while $I_{M_1} \longrightarrow P_1$ is productive.

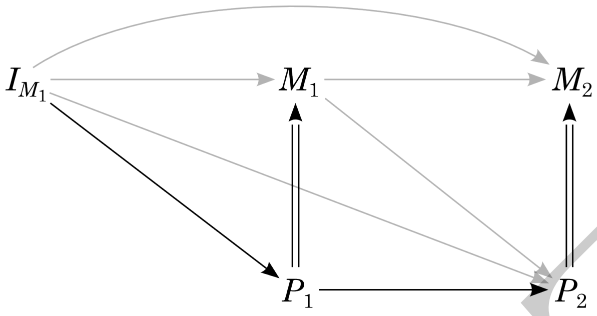

Figure 3: The grey arrows indicate (possible) direct causal relations that cannot propagate dependence between the variables at their heads and tails. So P_1 is the only variable in $\{M_1, M_2, P_1, P_2\}$ that can be directly intervened on in a productive way.

Prod (which seems to be reasonable when aiming at finding empirical evidence for causal relations $M_1 \longrightarrow M_2$ and $M_1 \longrightarrow P_2$), it is impossible to test for whether the arrows $M_1 \longrightarrow P_2$ and $M_1 \longrightarrow M_2$ are productive by means of interventions, even when allowing for common cause interventions I_{M_1} of M_1 and P_1.

Note that I_{M_1}, though inefficacious w.r.t. M_1 over the arrow $I_{M_1} \longrightarrow M_1$, can still be understood as an intervention on M_1 (since I_{M_1} might influence M_1 over path $I_{M_1} \longrightarrow P_1 \Longrightarrow M_1$). So our findings still allow for a change in M_2 or P_2 induced by an intervention on M_1, and we can still interpret the experimental result that intervening on M_1 leads to a change in M_2 or P_2 as evidence that by intervening on M_1 we can bring about (or at least influence) M_2 or P_2, respecitvely. All the causal work, however, is done by the path $I_{M_1} \longrightarrow P_1 \longrightarrow P_2$, and we are not allowed to interpret changes in M_2 or P_2 induced by an intervention $I_{M_1} = i_{M_1}$ as evidence for the presence of an efficacious direct causal connection $M_1 \longrightarrow M_2$ or $M_1 \longrightarrow P_2$, respectively. This will, of course, not distress scientists doing experiments too much, since whether M_1 or P_1 does all the causal work will not make any difference for the experiment's outcome.

Summarizing, our findings strengthen Baumgartner's (2013) results. It is not only the case that until now we do not know how to find empirical evidence for M_1's causal efficacy on P_2 or M_2 within an interventionist framework; rather it seems generally (or theoretically) impossible that M_1 has a causal influence on P_2 or M_2. In addition, a common cause (or fat-handed) intervention I_{M_1} for M_1 and P_1 cannot directly influence M_1. This means that attempts to render the causal effectiveness of mental properties on physical properties or on other mental properties plausible on empirical grounds within an interventionist framework seemed to be deemed to failure *ab initio*.

5 Woodward's objection to treating supervenience relations like causal arrows

In this section I defend the suggestion to treat supervenience relationships like causal arrows in a CBN (for which I argued in section 3) against an objection raised by Woodward (2014). Woodward's objection is that treating supervenience like a causal relation would lead to absurd consequences and contradicts experimental practice. Woodward (2014, sec. 6) comes up with the following example to demonstrate this: Assume that high density cholesterol (HDC) and low density cholesterol (LDC) are both causes of having a heart disease (D). While high density cholesterol lowers the probability of heart disease, low density cholesterol raises the probability for heart disease. Let TC be a variable for total cholesterol that is defined as $TC = HDC + LDC$. Hence, TC will supervene on HDC and LDC. Now Woodward assumes that HDC, LDC, and TC are causes of D. If we want to represent all of these variables in a single CBN, this

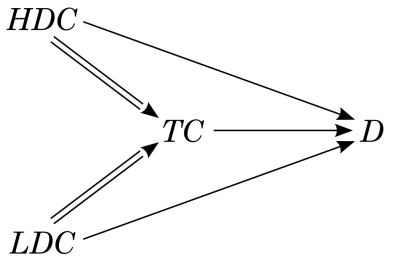

Figure 4: A CBN including D, TC, and TC's supervenience base. Double-tailed arrows, again, stand for supervenience relations.

CBN's graph would—following my suggestion to treat supervenience relations like causal relations—look like the one in Figure 4. Again, double-tailed arrows are assumed to technically work exactly like single-tailed arrows.

Now Woodward's (2014) objection against treating double-tailed arrows like causal arrows roughly goes as follows: For testing whether LDC is a direct (and efficacious) cause of D, one has to fix all remainder variables by means of interventions and check whether intervening on LDC leads to a change in D. But when we run this test, then, since TC is defined as $TC = HDC + LDC$, we are not able to manipulate LDC when HDC and TC are both fixed by interventions. Thus, the interventionist theory of causation would tell us that LDC has no effect on D, and, moreover, that LDC is not even a cause of D. Similarly it can be shown that neither HDC nor TC would have an effect on D and that neither HDC nor TC would turn out to be causes of D within the interventionist framework. All of these consequences contradict the assumptions made above. Furthermore, treating supervenience relationships similar to causal arrows does not conform to experimental practice. No researcher would seriously consider to fix TC's supervenience base LDC and HDC when testing TC's causal efficacy w.r.t. D. According to Woodward, this would amount to double counting the effect of TC on D. Similar worries apply w.r.t. LDC and HDC.

Woodward (2014) interprets this observation as support for his claim that double-tailed arrows standing for supervenience relationships should not be treated like causal arrows and as motivation for modifying his interventionist theory of causation in such a way that it is no longer required to hold fixed variables stnading in non-causal relationships (such as supervenience relationships) when testing for causal dependence and efficacy. But does Woodward's observation really threaten my suggestion to treat supervenience relations like causal arrows in CBNs? I will argue that this is not the case. Actually, the problems Woodward describes arise only within an interventionist framework, but not within the CBN framework. Let me illustrate this by reconstructing the scenario described in the first paragraph of this section as a CBN. This CBN's graph would, of course, again be the one depicted in Figure 4. Since the possibility to intervene on LDC and induce changes on D by means of this intervention when fixing HDC and TC by additional interventions is not required within the CBN framework for direct causation, we do not have to infer that LDC is not a direct cause of D. The same holds for HDC and TC. So we can avoid this problem.

The second problem Woodward (2014) sees is that neither LDC, nor HDC or TC would turn out as efficacious w.r.t. D. Can we also avoid this problem? We can test each of the causal arrows $LDC \longrightarrow D$, $HDC \longrightarrow D$, and $TC \longrightarrow D$ for productiveness in our CBN. If we do this, we find—by using the argumentation pattern described in section 3—that none of these arrows is productive, simply because every variable's value is fully determined by the values of the other two variables. What should we make of this observation? First of all, let me emphasize that this situation is not so special that it can only occur in the presence of supervenience relationships. One can easily construct an equivalent (purely) causal model with different variables but with the same topological structure and the same dependencies in which the double-tailed arrows are replaced by ordinary single-tailed causal arrows. More generally, the productivity test suggested in section 3 tells us that all of a variable's causes are causally inafficacious if every one of these direct causes is fully determined by the other direct causes. What the productivity test would indicate in such a situation is that at least one of the causal arrows should be deleted. (Recall from section 3 that the productivity condition is equivalent with the causal minimality condition.) After deleting one arrow, the productiveness of the remaining arrows may be restored. The same holds for our CBN.

So which arrow(s) should we delete? If we delete $LDC \longrightarrow D$, then $HDC \longrightarrow D$ and $TC \longrightarrow D$ become productive in the resulting model. If we delete $HDC \longrightarrow D$, then $LDC \longrightarrow D$ and $TC \longrightarrow D$ become productive. And if we delete $TC \longrightarrow D$, then $LDC \longrightarrow D$ and $HDC \longrightarrow D$ become productive in the resulting model. Every one of these possible deletions of arrows would result in a causal model that still satisfies the causal Markov condition. If we delete more than one arrow, then the causal Markov condition would be violated (since the resulting graph would imply more probabilistic independencies than featured by our example). So, to account for all the (conditional and unconditional) dependencies among our four variables, we should only delete one arrow. Now we have two possibilities: We delete (i) one of the arrows exiting one of the variables of TC's supervenience base, or we delete (ii) the arrow $TC \longrightarrow D$. If we decide in favor of (i), then we could ask ourselves why we should delete $LDC \longrightarrow D$ rather than $HDC \longrightarrow D$ (or vice versa). Which one of the two arrows we delete seems quite arbitrary. In addition, it would be strange to assume that the macro property TC and only one of its constituting properties is causally efficacious, while the other one is not. So it seems much more natural to decide in favor of (ii) and delete the arrow $TC \longrightarrow D$ instead. If we do this, then the resulting CBN gives us everything Woodward (2014) requested except that TC is causally efficacious w.r.t. D: LDC and HDC are causally efficacious w.r.t. D, and also TC-changes are associated with D-changes, simply because TC is constituted by LDC and HDC.

Our CBN even mirrors scientific practice and provides the correct results about the effects of interventions: An intervention on LDC, for example, can be modeled by adding an intervention variable I_{LDC}, which is a direct cause only of LDC. Intervening on LDC corresponds to conditionalizing on one of I_{LDC}'s on-values and will have an effect on D. This effect solely arises due to the path $I_{LDC} \longrightarrow LDC \longrightarrow D$. There is no double counting involved here. The same holds for an intervention on HDC: Every effect of an intervention $I_{HDC} = i_{HDC}$ on D will solely arise due to the causal path $I_{HDC} \longrightarrow HDC \longrightarrow D$. We can also handle an intervention on the constituted variable TC. Such an intervention could be represented by an intervention variable I_{TC}. We add I_{TC} as a direct cause of TC and assume that TC can be influenced by I_{TC}. By means of the same argumentation pattern applied earlier in such situations, it turns out that the arrow $I_{TC} \longrightarrow TC$ is unproductive, since TC is constituted by LDC and HDC and, hence, determined by LDC and HDC. It follows that I_{TC} must influence TC over another path. The only possibility to do so is over one of the paths $I_{TC} \longrightarrow LDC \Longrightarrow TC$ or $I_{TC} \longrightarrow HDC \Longrightarrow TC$. So I_{TC} has to be a common cause of TC and at least one of the variables LDC or HDC. Now a change of I_{TC}'s value may, of course, not only influence TC over one of these paths, but also D over one of the paths $I_{TC} \longrightarrow LDC \longrightarrow D$ or $I_{TC} \longrightarrow HDC \longrightarrow D$. So, again, we get everything Woodward requested except the productive causal arrow $TC \longrightarrow D$. We can even say that an intervention on TC leads to a change in D, which might be something we find out in doing an experiment. But as in the case of the causal exclusion scenario, we should

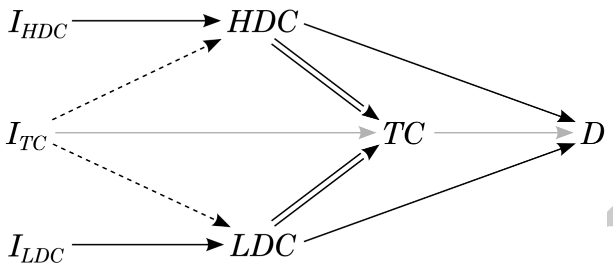

Figure 5: Grey arrows stand for (possible) direct causal relations that cannot propagate dependence between the variables at their heads and tails. I_{TC} is a fat-handed intervention, i.e., a common cause of TC and at least one of the variables LDC or HDC. This is indicated by the dashed arrows.

not read this as evidence for TC being causally efficacious w.r.t. D. The whole causal work is done by one or both of the causal paths $I_{TC} \longrightarrow LDC \longrightarrow D$ and $I_{TC} \longrightarrow HDC \longrightarrow D$. But again, this should not be too cumbersome for the scientist. Whether the intervention is directly efficacious w.r.t. TC or effects TC only over one or both of its supervenience bases will not be of much interest to her, simply because it will not make any difference for the experiment's outcome. Note that also testing whether D can be influenced by manipulating TC does not involve any double counting.

Summarizing, the problems Woodward (2014) describes arise only within an interventionist framework and not when treating supervenience relationships like causal arrows in a CBN, as I have suggested in section 3. These findings are graphically illustrated in Figure 5. There may be other objections against my suggestion to treat supervenience relationships like causal arrows in CBNs. But as far as I can see such objections still wait to be discovered and formulated.

6 Conclusion

In this paper I reconstructed two variants of the causal exclusion argument within the theory of CBNs. This seems promising since the theory of CBNs probably gives us the best grasp on causation from an empirical point of view we have so far. The reconstruction required to represent Kim's (2005) diagram as a CBN. Causal relations in Kim's diagram can straightforwardly be represented by a CBN's causal arrows. I argued that since relationships of supervenience behave exactly like causal arrows in CBNs, the double-tailed arrows standing for such relationships can be treated like ordinary single-tailed causal arrows in a CBN. The CBN's probability distribution is constrained by the assumptions that P_1 fully determines P_2 (completeness of the physical), that every change of M_i's value leads to a probability change of at least one P_i-value (supervenience), and that M_i is fully determined by P_i (constitution). For this CBN it turned out that both causal arrows $M_1 \longrightarrow P_2$ and $M_1 \longrightarrow M_2$ are unproductive, meaning that they cannot transport probabilistic dependence. Because of the very nature of how P_2 depends on P_1 (completeness of the physical) and of how M_i depends on P_i (constitution), this result generalizes to all expansions of the CBN. Thus, both variants of the exclusion argument are valid under the proviso that causes contribute at least sometimes something to the occurrence of their effects.

In section 4 I discussed the consequences of these findings for the discussion of causal exclusion arguments in the light of an interventionist theory of causation. Our findings strengthen Baumgartner's (2013) criticism of Woodward (2014). Baumgartner concludes that it is unclear how one could provide empirical evidence for a mental property's causal efficacy on physical properties within an interventionist framework. We could show that such a mental property M_1's causal efficacy on a physical property P_2 or on another mental property M_2 cannot be empirically supported at all, simply because causal arrows $M_1 \longrightarrow P_2$

and $M_1 \longrightarrow M_2$ are always unproductive, meaning that they do not imply any correlation between M_1 and P_2 and between M_1 and M_2, respectively, in any circumstances. Moreover, it could be shown that it is generally impossible to have a causally productive direct intervention on a mental property, and thus, that attempts to investigate whether mental properties can be causally efficacious within an interventionist framework were somehow unlucky from the beginning.

In the last section of this paper I discussed an objection against modeling supervenience relationships similar to causal arrows raised by Woodward (2014). I argued that Woodward's objection does not pose a threat to my suggestion of how to represent supervenience in CBNs. His objection works only within an interventionist framework. The problems he highlights do not appear in CBNs in which double-tailed arrows indicating supervenience relationships are treated similar to single-tailed arrows standing for direct causal dependencies.

Acknowledgements: This work was supported by Deutsche Forschungsgemeinschaft (DFG), research unit FOR 1063. My thanks go to Christopher R. Hitchcock, Andreas Hüttemann, Markus Schrenk, and Gerhard Schurz for important discussions. Thanks also to Alexander G. Mirnig, Christian J. Feldbacher, Wlodek Rabinowicz, and an anonymous referee for helpful comments on earlier versions of this paper.

References

Baumgartner, M. (2009). Interventionist causal exclusion and non-reductive physicalism. *International Studies in the Philosophy of Science*, *23*(2), 161–178.

Baumgartner, M. (2010). Interventionism and epiphenomenalism. *Canadian Journal of Philosophy*, *40*(3), 359–383.

Baumgartner, M. (2013). Rendering interventionism and non-reductive physicalism compatible. *Dialectica*, *67*(1), 1–27.

Baumgartner, M., & Gebharter, A. (2015, February). Constitutive Relevance, Mutual Manipulability, and Fat-Handedness. *British Journal for the Philosophy of Science*, axv003.

Eronen, M. I. (2012). Pluralistic physicalism and the causal exclusion argument. *European Journal for Philosophy of Science*, *2*(2), 219–232.

Gebharter, A., & Schurz, G. (2014). How Occam's razor provides a neat definition of direct causation. In J. M. Mooij, D. Janzing, J. Peters, T. Claassen, & A. Hyttinen (Eds.), *Proceedings of the uai workshop causal inference: Learning and prediction.* Aachen.

Harbecke, J. (2013). On the distinction between cause-cause exclusion and cause-supervenience exclusion. *Philosophical Papers*, *42*(2), 209–238.

Hitchcock, C. (2012). Theories of causation and the causal exclusion argument. *Journal of Consciousness Studies*, *19*(5-6), 40–56.

Kim, J. (1989). Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives*, *3*, 77–108.

Kim, J. (2000). *Mind in a physical world.* MIT Press.

Kim, J. (2003). Blocking causal drainage and other maintenance chores with mental causation. *Philosophy and Phenomenological Research*, *67*(1), 151–176.

Kim, J. (2005). *Physicalism, or something near enough.* Princeton University Press.

McLaughlin, B., & Bennett, K. (2011). Supervenience. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy.*

Neapolitan, R. E. (1990). *Probabilistic reasoning in expert systems.* Wiley.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Pearl, J. (2000). *Causality* (1st ed.). Cambridge: Cambridge University Press.

Raatikainen, P. (2010). Causation, exclusion, and the special sciences. *Erkenntnis*, *73*(3), 349–363.

Robinson, W. (2015). Epiphenomenalism. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy.*

Schurz, G. (2008). Patterns of abduction. *Synthese*, *164*(2), 201–234.

Schurz, G., & Gebharter, A. (2015). Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets. *Synthese*.

Shapiro, L. A. (2010). Lessons from causal exclusion. *Philosophy and Phenomenological Research*, *81*(3), 594–604.

Shapiro, L. A., & Sober, E. (2007). Epiphenomenalism – the Do's and the Don'ts. In G. Wolters & P. Machamer (Eds.), *Studies in causality: Historical and contemporary* (pp. 235–264).

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

Woodward, J. (2003). *Making things happen.* Oxford: Oxford University Press.

Woodward, J. (2008). Mental causation and neural mechanisms. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced: New essays on reduction, explanation, and causation* (pp. 218–262). Being Reduced.

Woodward, J. (2014). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*.

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