Causal exclusion without physical completeness and no overdetermination[*]

Alexander Gebharter

Abstract: Hitchcock (2012) demonstrated that the validity of causal exclusion arguments as well as the plausibility of several of their premises hinges on the specific theory of causation endorsed. In this paper I show that the validity of causal exclusion arguments—if represented within the theory of causal Bayes nets the way Gebharter (2015) suggests—actually requires much weaker premises than the ones which are typically assumed. In particular, neither completeness of the physical domain nor the no overdetermination assumption are required.

1 Introduction

Causal exclusion arguments (cf. Kim, 2000, 2005) are typically used as arguments against non-reductive physicalism or as arguments for epiphenomenalism. They conclude from several premises that mental properties cannot be causally

[*] This is a draft paper. The final version of this paper is published under the following bibliographical data: Gebharter, A. (2017). Causal exclusion without physical completeness and no overdetermination. *Abstracta – Linguagem, Mente e Ação, 10*, 3–14.

efficacious. The premises typically endorsed are the following (cf. Woodward, 2015, sec. 2; Hitchcock, 2012, pp. 42ff):

> **Distinctness:** Mental properties cannot be reduced to physical properties; they are ontologically distinct.
>
> **Supervenience:** Mental properties supervene on physical properties.[1]
>
> **Physical completeness:** Every physical property has a sufficient physical cause.[2]
>
> **No overdetermination:** No property has more than one sufficient cause.

In a nutshell, exclusion arguments run as follows: Let M be a mental property and let P be M's physical supervenience base. Now assume X to be a spatio-temporally distinct (mental or physical) property. Let us further assume that all three properties are instantiated. In case X is a mental property, X has a supervenience base Y which is also instantiated and fully determines X. In that case, X is instantiated because Y is instantiated and there is nothing left M could contribute to whether X occurs. In case X is a physical property, there is a sufficient physical cause Y of X. This sufficient physical cause Y is either P alone or P together with some other physical cause(s) of X. Also in

[1] Supervenience is understood as strong supervenience here, meaning that every change in the supervening property is necessarily accompanied by a change in its supervenience base, while the supervenience base determines the supervening property (with probability 1).

[2] There are also weaker versions of the physical completeness principle which say that every physical effect has a sufficient physical cause. The difference between the two is, however, not that important for most of what I will do in this paper. Hence, I will most of the time stick to physical completeness as introduced here.

that case X is instantiated because Y is instantiated; there is nothing left M could contribute to whether X occurs. Since M and X were arbitrarily chosen, the argument generalizes: There is no mental property M and no property X spatio-temporally distinct from M and its supervenience base such that M can contribute anything to whether X occurs. Hence, mental properties are causally inefficacious.

Hitchcock (2012) convincingly demonstrated that the validity of causal exclusion arguments as well as the plausibility of several of their premises hinges on the specific theory of causation endorsed. In particular, he showed that for three different theories of causation, *viz.* Laplacean causation, process causation, and difference-making causation, at least one of the premises mentioned above is not plausible. Gebharter (2015) provided a reconstruction of causal exclusion arguments within another theory of causation, *viz.* the theory of causal Bayes nets (CBNs), and proved their validity (given the reconstruction of supervenience relationships he suggested is correct). He did, however, not say anything about the status of the premises typically used in such arguments within the CBN framework. This is what I will do in this paper. After briefly introducing some basics of the theory of CBNs and presenting the reconstruction of causal exclusion arguments suggested in (Gebharter, 2015) (section 2), I argue that physical completeness as well as the no overdetermination assumption, which have some weak spots which could be atacked from friends of non-reductive physicalism, are not required for the argument to go through (section 3). One nicely gets the conclusion of causal exclusion arguments within a CBN framework by assuming instead the quite harmless principle that if mental properties are causally efficacious, then also their physical supervenience bases are. This result strenghtens exclusion arguments as arguments against non-reductive physicalism and as evidence for epiphenomenalism from the perspective of a CBN framework. I conclude in section 4.

2 Causal exclusion and causal Bayes nets

A CBN is a triple $\langle V, E, P \rangle$. V is a set of random variables, $G = \langle V, E \rangle$ is a directed acyclic graph, and P is a probability distribution over V. E is a set of directed edges ($\longrightarrow$) between variables in V. G's edges $X \longrightarrow Y$ are interpreted as direct causal relations w.r.t. V. The variables X at the ends of the arrows pointing at another variable Y in G are called Y's parents ($Par(Y)$). The variables Y which are connected to another variable X via a chain of arrows of the form $X \longrightarrow ... \longrightarrow Y$ are called X's descendants ($Des(X)$). CBNs are assumed to satisfy the causal Markov condition (CMC) (Spirtes, Glymour, & Scheines, 2000, p. 29):

Definition 2.1 (causal Markov condition)**.** *$\langle V, E, P \rangle$ satisfies the causal Markov condition if and only if $Indep(X, V \backslash Des(X) | Par(X))$ holds for all $X \in V$.*[3]

CMC generalizes the Reichenbachian insight that conditionalizing on all common causes renders two formerly correlated variables independent, while conditionalizing on a variable's direct causes renders it independent of its indirect causes (cf. Reichenbach, 1956/1991). It lies at the very heart of the theory of causal Bayes nets and establishes an intimate connection between unobservable (theoretical) causal structures and empirically accessible probability distributions. It plays an important role for formal causal reasoning, for formulating and testing of causal hypotheses, (together with other conditions)

[3] $Indep(X, Y|Z)$ stands for probabilistic independence of X on Y conditional on Z, which is defined as $P(x|y,z) = P(x|z) \vee P(y,z) = 0$ for all x, y, z. $Dep(X, Y|Z)$ stands short for dependence of X on Y given Z, which is defined as the negation of $Indep(X, Y|Z)$, i.e., as $P(x|y,z) \neq P(x|z) \wedge P(y,z) > 0$ for some x, y, z.

for causal discovery, and for computing the effects of interventions even if only non-experimental data is available (see, e.g., Spirtes et al., 2000).

Whenever CMC is satisfied, our CBN's graph determines the following Markov factorization (cf. Pearl, 2000, sec. 1.2.2):

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

Basically all kinds of relations that produce the Markov factorization can be represented by the arrows of a CBN. Direct causation is only one of these relations. Gebharter (2015) argued that supervenience is another such relation. Whether this argumentation is correct is still debatable. For this paper, however, I will take it for granted that supervenience can be represented like direct causal connection within CBNs. Or in other words: The present paper investigates which typical premises of causal exclusion arguments are actually needed if the argumentation provided by Gebharter is correct. If it is correct, then direct causation as well as supervenience can be represented by the arrows of a CBN.[4] (Note that I do not want to claim that supervenience is a special form of causation; I prefer to stay neutral on this ontological question.) In the following, we will represent direct causal relations by means of single-tailed arrows, and relationships of supervenience by means of double-tailed arrows. Both kinds of arrows are assumed to technically work like ordinary single-tailed causal arrows in a CBN.

Gebharter (2015) reconstructs causal exclusion arguments with help of the

[4] Many philosophers seem to think that also another condition, *viz.* the faithfulness condition (see Spirtes et al., 2000, p. 31), has to be satisfied. This is, however, not true. Faithfulness is a nice thing to have for many reasons, first and foremost it is essential for causal discovery. Faithfulness is, however, not necessary for representing a system's causal structure by means of a CBN. Everything needed for a CBN is that the Markov condition is satisfied.

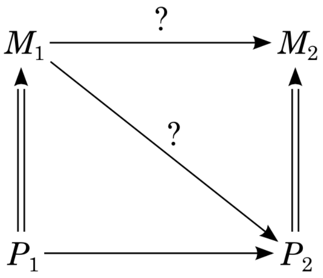

Figure 1

CBN depicted in Figure 1. M_1, M_2 stand for mental properties, and P_1, P_2 stand for their respective physical supervenience bases. It is assumed that P_1 is P_2's sufficient physical cause. The question marks over the arrows $M_1 \longrightarrow M_2$ and $M_1 \longrightarrow P_2$ indicate that these two arrows are the ones which should be tested for causal effectiveness.

Note that the theory of CBNs comes with the following neat test for whether particular causal arrows can produce probabilistic dependence: To test for whether $X \longrightarrow Y$ is productive, check whether $Dep(Y, X|Par(Y)\backslash\{X\})$ holds (cf. Gebharter, 2015; Schurz & Gebharter, 2016). If yes, then $X \longrightarrow Y$ is productive. If no, then X cannot have a direct causal influence on Y. Informally speaking, we test for whether X can have an influence on its direct effect Y in any circumstances, i.e., in the light of any causal background story. When this test is applied to the causal exclusion CBN, it turns out that both arrows $M_1 \longrightarrow M_2$ and $M_1 \longrightarrow P_2$ are unproductive, meaning that M_1 is causally inefficacious w.r.t. both M_2 and P_2.

In particular, the argumentation for the unproductiveness of the arrow $M_1 \longrightarrow M_2$ runs as follows (Gebharter, 2015, sec. 3): Let p_2 be an arbitrarily chosen P_2-value. Recall that M_2 supervenes on P_2. This implies that M_2's value is fully determined by P_2's value, i.e., that 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 $m_2' \neq m_2$. 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. It is probabilistically valid that conditional probabilities of 1 and 0 cannot be changed when conditionalizing on compatible values of additional variables. Because of this, $P(m_2|m_1, p_2) = P(m_2|p_2) = 1$ and $P(m_2'|m_1, p_2) = P(m_2'|p_2) = 0$ 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 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, the result obtained in both cases can be generalized: 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 $Par(M_2)\backslash\{M_1\} = \{P_2\}$. It not follows directly from the definition of productivity that the arrow $M_1 \longrightarrow M_2$ is unproductive.

The argumentation for the unproductiveness of the arrow $M_1 \longrightarrow P_2$ runs as follows (Gebharter, 2015, sec. 3): Let p_1 be an arbitrarily chosen P_1-value. Because P_1 is assumed to be P_2's sufficient cause, P_2's value is fully determined by P_1's value. Because of this for every p_1 there is exactly one p_2 such that $P(p_2|p_1) = 1$, while $P(p_2'|p_1) = 0$ 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., $P(m_1, p_1) > 0$. Since conditionalizing on compatible values of additional variables cannot have any influence on conditional probabilities of 1 and 0, conditionalizing on m_1 will not change the conditional probabilities of p_2 or p_2' given p_1, i.e., also $P(p_2|m_1, p_1) = P(p_2|p_1) = 1$ and $P(p_2'|m_1, p_1) = P(p_2'|p_1) = 0$ 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., $P(m_1, p_1) = 0$. It then follows, again from the definition of probabilistic independence, 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, the result obtained in the two cases can be generalized: 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 $Par(P_2)\backslash\{M_1\} = \{P_1\}$. From our productivity test it follows then that the arrow $M_1 \longrightarrow P_2$ is unproductive.

3 Physical completeness and no overdetermination within the CBN framework

Gebharter's (2015, sec. 3) reconstruction of the exclusion argument seems to make use of all four premises introduced in section 1. Because of the distinctness premise, mental properties are represented by different variables (M_1, M_2) than the ones (P_1, P_2) used to represent their respective physical supervenience bases. The supervenience premise implies some constraints on the CBN's probability distribution, *viz.* that every change in M_i's value leads to a probability change of some P_i-value and that every P_i-value determines M_i to take a specific value with probability 1. The premise of the completeness of the physical domain implies that for every physical property represented by a variable P_i there is a sufficient physical cause, i.e., a variable P_j such that P_i is fully determined by P_j. The CBN reconstruction assumes P_1 to be such a sufficient physical cause of P_2. Finally, the no overdetermination assumption seems to be present in the productivity test applied to the causal arrows $M_1 \longrightarrow M_2$ and $M_1 \longrightarrow P_2$: M_1 is only accepted as causally efficacious if there is no systematic overdetermination, i.e., if M_1 has at least a slight influence on M_2's or on P_2's probability distribution when all parents of M_2 different from M_1 or all parents of P_2 different from M_1 are fixed to certain values.

The majority of philosophers and philosophically minded scientists seems to accept that mental properties supervene on physical properties. Every change of a decision, for example, is necessarily accompanied by changes in the brain and also fully determined (or constituted) by these changes. So the supervenience premise seems to be quite harmless and basically everyone wants to subscribe to it. Concerning the distinctness premise, I have neither any evidence for nor any intuition about whether it is true. However, if mental properties are not distinct from physical properties, then there seems to be little space for them to be autonomous in the sense the non-reductive physicalist would like them to be. And if mental properties are distinct from physical properties, then non-reductive physicalism seems to fall prey to the exclusion argument (at least within the theory of CBNs). Either way this is bad news for the supporter of non-reductive physicalism. To give non-reductive physicalism a chance, however, one has to assume distinctness. For the reasons mentioned I will leave the distinctness assumption and the supervenience premise untouched and will not discuss them in more detail in the remainder of this paper. I will rather focus on the more interesting premises which also clearly refer to causation: the physical completeness premise and the no overdetermination premise.

Let us start with a closer look at the assumption of the completeness of the physical domain. Though this premise is in principle compatible with the theory of CBNs, there are several possibilities for the non-reductive physicalist to attack it. One worry the non-reductive physicalist might have is, for example, that physical completeness is a quite strong metaphysical assumption. Why should we believe that really every physical property has a sufficient physical cause? The big bang, for example, might be an uncaused event. There are, however, weaker versions of the physical completeness premise available on the market which can avoid this worry. One might, for example, only assume that there is a sufficient physical cause for every caused physical event (cf. Esfeld, 2007; Papineau, 1993). This version of physical completeness would clearly allow for uncaused events like the big bang. And it would still be sufficient to run the exclusion argument within the CBN framework. If M_1 causes P_2, then P_1 is a sufficient cause of P_2 and there is no causal role left over for M_1 to play.[5] But also this version as a premise seems to be quite strong. It excludes events which are only caused in a purely probabilistic way. An obvious example is the decay of uranium, which can only be probabilistically influenced. But if we have good reasons to doubt that every caused physical property has a sufficient physical cause, then Gebharter's (2015, sec. 3) argumentation for the unproductiveness of the arrow $M_1 \longrightarrow P_2$ does not go through. If it cannot be guaranteed that P_1 fully determines P_2, then—so it seems—it might happen that P_2 still depends on M_1 when conditionalizing on P_1. In that case, the productivity test would tell us that M_1 can be causally efficacious w.r.t. P_2 and that non-reductive physicalism could—at least in principle—be saved.

I agree that Gebharter's (2015, sec. 3) original argument for the unproductiveness of the arrow $M_1 \longrightarrow P_2$ would be undermined if we are not allowed to assume that P_1 fully determines P_2 anymore. However, there is a slightly different argument for the unproductiveness of this particular arrow that does not require P_1 to be a sufficient cause of P_2. It only requires the following as a premise instead:

[5] Note that the argumentation for the unproductiveness of the arrow $M_1 \longrightarrow P_2$ does not depend on a causal relation between P_1 and M_2 at all. For showing that this arrow is unproductive, physical completeness is, hence, not required.

> **No mental causation without physical causation:** If a mental property M is a cause of a physical property X, then also M's physical supervenience base P is a cause of X.

This assumption is weaker than the two versions of the assumption of the completeness of the physical domain mentioned above. The stronger one of the two versions of the physical completeness premise leads to infinitely many physical events in one's ontology once there is at least one such physical event: If there is a physical event e_1, then there is also e_1's sufficient physical cause e_2. But e_2's existence requires another sufficient physical cause e_3 and so on ad infinitum. On the other hand, the no mental causation without physical causation principle stated above neither requires that all physical events are caused, nor that there are any sufficient physical causes at all. It just says that *if* there is a mental property that causes some physical property X, *then* also this mental property's supervenience base is causally relevant for X (in a deterministic or an indeterministic way). This seems to be a highly plausible assumption. It is clearly weaker than the stronger version of the premise of the completeness of the physical domain. From the pure existence of a physical event e_1 (alone) nothing follows according to the no mental causation without physical causation principle. The existence of other physical causes only follows if there are also mental causes of e_1. And even in that case these additional physical causes might be weak indeterterministic causes. Hence, the no mental causation without physical causation principle is also weaker than the weaker one of the two versions of the physical completeness premise, which only requires that caused physical events have sufficient physical causes.

Now one might think that the no mental causation without physical causation principle is, in truth, just a weaker version of the physical completeness premise. I think that the former is not just a weaker version of the latter. There is another crucial difference between the two assumptions. The mental causation without physical causation principle connects mental causation to physical causation. It says that certain phsical causal facts have to hold *if* certain mental causal facts hold. For the reductive physicalist, the principle is empty, simply because she believes that mental facts are nothing over and above physical facts. For her the principle just says that properties which have physical causes have physical causes. The physical completeness premise, on the other hand, is not empty for the reductive physicalist. For her the physical completeness premise still implies the existence of sufficient physical causes if there are any (physical) causes.

Before we go on, let me briefly illustrate the no mental causation without physical causation principle by means of Hitchcock's (2012, p. 42) refrigerator example: I decide to go to the refrigerator to grab something to drink. The decision is the mental event, certain changes in my brain form its physical supervenience base, and my body moving toward the refrigerator is the physical event I intend to bring about. Now let us assume that my decision causes my body to move toward the refrigerator (in a deterministic or indeterministic way). In that case—without much doubt—also the changes in my brain on which my decision supervenes will be causally relevant for my body moving toward the refrigerator. Note how weak the no mental causation without physical causation principle actually is: In case epiphenomenalism or reductionism is true, there are no mental causes (different from brain processes) and, hence, the principle keeps silent about the existence of any physical causes of my body's moving toward the refrigerator different from mental properties. And even if there were mental causes—meaning that non-reductive physicalism were true—then the no mental causation without physical causation principle would only require that also these mental causes' physical supervenience bases are causes that make at least a slight probabilistic difference for my body's moving toward the refrigerator.

Now the assumption that there is no mental causation without physical causation is everything required to show that the arrow $M_1 \longrightarrow P_2$ is unproductive in the CBN depicted in Figure 1. In the original argument, the arrow $M_1 \longrightarrow P_2$ turned out as unproductive because P_2's parent P_1 was assumed to be a sufficient cause of P_2 and, hence, fully determined P_2's value. But if P_2's value is determined by P_1, then no change in M_1 can be associated with a change in P_2. Thus, we get the independence $Indep(P_2, M_1|P_1)$. But P_1 does not only determine P_2, but also M_1 (because M_1 supervenes on P_1). So we do not even need the arrow $P_1 \longrightarrow P_2$ to be deterministic, or, in other words: We do not even need P_1 to be a sufficient cause of P_2 to get the independence $Indep(P_2, M_1|P_1)$.

Here is the argument: Let p_1 be an arbitrarily chosen P_1-value. Due to the fact that M_1 supervenes on P_1, P_1 fully determines M_1. Hence, there is exactly one M_1-value m_1 for every P_1-value p_1 such that $P(m_1|p_1) = 1$ holds, while $P(m'_1|p_1) = 0$ holds for all $m'_1 \neq m_1$. Now for every single P_2-value p_2 there are two possible cases.

Case 1: p_1 and p_2 are compatible, i.e., $P(p_1, p_2) > 0$. Because conditionalizing on compatible values of additional variables cannot have any influence on conditional probabilities of 1 and 0, also $P(m_1|p_1, p_2) = P(m_1|p_1) = 1$ and $P(m'_1|p_1, p_2) = P(m'_1|p_1) = 0$ will hold. Hence, no M_1-value depends on p_2 conditional on p_1.

Case 2: p_1 and p_2 are incompatible, meaning that $P(p_1, p_2) = 0$. From this it follows by the definition of probabilistic independence that no M_1-value depends on p_2 conditional on p_1. Therefore, conditionalizing on p_1 renders p_2 probabilistically independent from M_1.

Again, p_1 was arbitrarily chosen for both cases above. Hence, the result obtained in both cases can be generalized: Conditionalizing on any P_1-value will render M_1 probabilistically independent from P_2. This is equivalent with $Indep(P_2, M_1|Par(P_2)\backslash\{M_1\} = \{P_1\})$. From $Indep(P_2, M_1|Par(P_2)\backslash\{M_1\} = \{P_1\})$ and our productivity test it follows that M_1 cannot have any probabilistic influence on P_2 over the arrow $M_1 \longrightarrow P_2$.

As a last step, let us also take a brief look at the plausibility and the role of the no overdetermination assumption within the CBN framework. Within this framework, the no overdetermination assumption basically corresponds to assuming the causal minimality condition (cf. Spirtes et al., 2000, p. 31), which is satisfied by a CBN if and only if every arrow of the CBN is productive (Gebharter & Schurz, 2014, theorem 1). First of all, note that assuming minimality is perfectly rational from a methodological point of view: We only want to assume causal relations that are at least in principle identifyable by their empirical (probabilistic) footprints. Nevertheless, a supporter of non-reductive physicalism may, again, object that assuming no overdetermination (or minimality) for all kinds of systems is much too strong from a metaphysical point of view. I agree that this is a strong metaphysical claim and that it is—at least in principle—possible that there are causal relations out there in the world which are systematically overdetermined. Let us grant this to the non-reductive physicalist and see what it implies for the reconstruction of the exclusion argument by means of the CBN depicted in Figure 1.

The interesting thing we can learn from the CBN reconstruction is that causal efficacy and the presence of a causal relation are two slightly different things. Supporters of the causal exclusion argument may be perfectly happy with direct causal relations between M_1 and M_2 as well as P_2 as long as M_1 can be shown to be inefficacious, i.e., as long as it can be shown that these relations cannot propagate any probabilistic dependence. And this is exactly what the reconstruction suggested by Gebharter (2015) shows. It does not require the no overdetermination premise (or the assumtion of minimality) at all. The productivity test porposed can be applied to every single arrow and it can be shown that the arrows $M_1 \longrightarrow M_2$ and $M_1 \longrightarrow P_2$ are unproductive. Whether we believe in no overdetermination and take the results of our productivity test as evidence to remove the arrows or do not care about overdetermination at all and leave the arrows intact: In any case M_1 can be shown to have no direct (probabilistic) influence on M_2 or P_2 in any circumstances. In other words: Even if M_1 actually is a cause of M_2 or P_2, it is necessarily an inefficacious cause. I think that even epiphenomenalists would be happy with this particular kind of mental causation (if it deserves to be called mental causation at all).

4 Conclusion

Causal exclusion arguments typically rest on four premises which I labeled distinctness, supervenience, physical completeness, and no overdetermination in section 1. While it is uncontested that mental properties supervene on physical properties, the distinctness of mental properties and physical properties is questionable. However, for the kind of autonomy of the mental the non-reductive physicalist demands it is essential to assume the latter. In this paper I focused on the remaining two premises (physical completeness and no overdetermination), whose plausibility depends on the specific theory of causation endorsed. I argued that both premises do not stand in conflict with the theory of CBNs, but that friends of non-reductive physicalism have good reasons to not accept these two conditions. In particular, both are quite strong from a metaphysical point of view. I then took a closer look at the role of these two premises within Gebharter's (2015) reconstruction of the exclusion argument. It could be shown that exclusion arguments go through with much weaker premises within a CBN framework. In particular, the no overdetermination assumption is not required at all, and the completeness of the physical domain can be replaced by a weaker and more plausible premise. This premise states that if a mental property causes a physical property, then also this mental property's physical supervenience base is causally relevant for that physical property.

All in all, the results of this paper can be seen as evidence against non-reductive physicalism from the view point of causal Bayes nets. To refute non-reductive physicalism it basically suffices to either reject that mental properties are distinct from physical properties, or to accept that mental properties supervene on physical properties and that if mental properties are causes of physical properties, then also their physical supervenience bases are. The two latter assumptions seem highly plausible.

Note that the results of this paper only hold for the reconstruction of causal exclusion arguments within the CBN framework suggested by Gebharter (2015). However, a reconstruction within the theory of CBNs seems promising for several reasons. The theory seems to give us the best grasp of causation we have so far. It allows for the development of powerful discovery algorithms, for testing causal hypotheses, and even for predicting the effects of possible interventions on the basis of purely observational data (Spirtes et al., 2000). The theory also behaves like a modern empirical theory of the sciences. Its core axioms can be justified by an inference to the best explanation of certain statistical phenomena and several versions of the theory can be shown to have empirical content by whose means they become testable on purely empirical grounds (cf. Schurz & Gebharter, 2016).

Acknowledgements: This work was supported by Deutsche Forschungsgemeinschaft (DFG), research unit Inductive Metaphysics (FOR 2495). I would like to thank Gerhard Schurz for important discussions. Thanks also to Christian J. Feldbacher-Escamilla and an anonymous referee for helpful comments on an earlier version of this paper.

References

Esfeld, M. (2007). Mental causation and the metaphysics of causation. *Erkenntnis*, *67*, 207–220.

Gebharter, A. (2015). Causal exclusion and causal Bayes nets. *Philosophy and Phenomenological Research*.

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.* Retrieved from http://ceur-ws.org/Vol-1274/uai2014ci_paper1.pdf

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

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

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

Papineau, D. (1993). *Philosophical naturalism.* Oxford: Blackwell.

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

Reichenbach, H. (1956/1991). *The direction of time.* Berkeley: University of California Press.

Schurz, G., & Gebharter, A. (2016). Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets. *Synthese*, *193*(4), 1073–1103.

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

Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, *91*(2), 303–347.