Philosophy of Science (2026), **93**, 170–180
doi:10.1017/psa.2025.10102

PSA PHILOSOPHY OF SCIENCE ASSOCIATION

PHILOSOPHY OF SCIENCE

ARTICLE

Causal Direction in Causal Bayes Nets

Reuben Stern and Benjamin Eva

Duke University, Durham, NC, USA
Corresponding author: Reuben Stern; reuben.stern@duke.edu

(Received 04 June 2024; revised 27 June 2025; accepted 01 July 2025; first published online 13 August 2025)

Abstract

Some authors maintain that we can use causal Bayes nets to infer whether $X \rightarrow Y$ or $X \leftarrow Y$ by consulting a probability distribution defined over some exogenous source of variation for X or Y. We raise a problem for this approach. Specifically, we point out that there are cases where an exogenous cause of X (E_x) has no probabilistic influence on Y no matter the direction of causation—namely, cases where $E_x \rightarrow X \rightarrow Y$ and $E_x \rightarrow X \leftarrow Y$ are probabilistically indistinguishable. We then assess the philosophical significance of this problem and discuss some potential solutions.

1. Introduction

You've discovered that X and Y are non-spuriously correlated and are thus sure that either X causes Y or that Y causes X. But you aren't sure which. How should you go about figuring this out?

It is *prima facie* attractive to maintain that we can infer the direction of causation between X and Y from the temporal order of X and Y (such that we infer that $X \rightarrow Y$ if X temporally precedes Y and infer that $Y \rightarrow X$ if Y temporally precedes X). But there are at least two reasons that we may not endorse this as a fully general strategy. First, it relies on the controversial claim that causes always temporally precede their effects.[1] Second, no matter what we say about the conceptual relationship between causation and time, there are contexts in which this strategy is not helpful because the variables under consideration aren't presented with a temporal ordering.

What to do? A number of authors working within the graphical approach to causal modeling maintain that we can tackle this problem by checking what happens when we identify some exogenous source of variation for one of the variables under

Both authors accept full and equal responsibility for what follows.

[1]It is contestable whether causes always temporally precede their effects since (i) there may be good reason to countenance the possiblity of retrocausality (see, e.g., Price (1996)) and (ii) it may be overly restrictive to limit the domain of the causal relation to temporally ordered variables.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Philosophy of Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

consideration.[2] According to this line of reasoning, if an exogenous cause of X has an effect on Y, then we can infer that $X \rightarrow Y$, whereas if an exogenous cause of X has no effect on Y, then we can infer that $Y \rightarrow X$.

In what follows, we raise a problem for this approach to inferring the direction of causation. Specifically, we point out that there are cases where an exogenous cause of X (E_X) has no probabilistic influence on Y no matter which way the arrow of causation points—namely, cases where $E_X \rightarrow X \rightarrow Y$ and $E_X \rightarrow X \leftarrow Y$ are probabilistically indistinguishable. We then assess the philosophical significance of the problem and survey multiple ways that we might try to persist in maintaining that the arrow of causation is somehow grounded in probabilistic (in)dependencies.

2. Unidentifiable colliders and intransitive chains

The justification for inferring causal direction from exogenous variation often goes by way of an axiom that characterizes causal Bayes nets—namely, the Causal Markov Condition.[3]

> Causal Markov Condition (CMC): Given a causal graph G over variable set $\mathbf{V}$ and probability distribution $\mathbb{P}$ over $\mathbf{V}$, G and $\mathbb{P}$ satisfy the Causal Markov Condition if and only if any variable X in $\mathbf{V}$ is probabilistically independent of its non-descendants given its parents.[4]

The CMC implies that $E_X \rightarrow X \rightarrow Y$ is compatible with the unconditional probabilistic dependence of E_X and Y, but that $E_X \rightarrow X \leftarrow Y$ is incompatible with the unconditional probabilistic dependence of E_X and Y. Thus we can conclude that $E_X \rightarrow X \rightarrow Y$ when we are confronted with our inference problem and observe a probabilistic dependence between the exogenous cause E_X and Y. But while the CMC successfully licenses this inference, it does not by itself license us to infer that $Y \rightarrow X$ when an exogenous cause of X has no probabilistic effect on Y. This is because the CMC is in the business of saying what causal dependencies are implied by what probabilistic dependencies (or, contrapositively, what probabilistic independencies are implied by what causal independencies), rather than what probabilistic dependencies are implied by what causal dependencies. As far as the CMC is concerned, then, the unconditional probabilistic independence of E_X and Y is compatible with both $E_X \rightarrow X \rightarrow Y$ and $E_X \rightarrow X \leftarrow Y$.

In order to rule out that $X \rightarrow Y$ on the grounds that some exogenous cause of X has no probabilistic effect on Y, we must additionally help ourselves to some condition that licenses us to infer that E_X and Y are unconditionally probabilistically dependent (or correlated)[5] when $E_X \rightarrow X \rightarrow Y$. The much discussed Causal Faithfulness Condition (CFC) is up to the task since it says that a causal graph G over $\mathbf{V}$ is compatible with a

[2] Interventionist approaches to this problem arguably fall under this general umbrella, since the intervention on X is often treated as an exogenous cause of X. See Hausman and Woodward (1999) and Woodward (2003).

[3] See, e.g., Spirtes, Glymour, and Scheines (2000).

[4] We use "causal graph" to denote a directed acyclic graph whose directed edges (or arrows) admit to a causal interpretation.

[5] Here, and in what follows, we often speak of probabilistic dependence in terms of correlation.

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

Table 1. Example probability distribution

V_1	V_2	V_3	P
0	0	0	$\frac{2}{66}$
0	0	1	$\frac{1}{66}$
0	1	0	$\frac{1}{66}$
0	1	1	$\frac{9}{66}$
0	2	0	$\frac{2}{66}$
0	2	1	$\frac{18}{66}$
1	0	0	$\frac{2}{66}$
1	0	1	$\frac{1}{66}$
1	1	0	$\frac{2}{66}$
1	1	1	$\frac{18}{66}$
1	2	0	$\frac{1}{66}$
1	2	1	$\frac{9}{66}$

probability distribution $\mathbb{P}$ over $\mathbf{V}$ if and only if there are no conditional independencies in $\mathbb{P}$ that are not entailed by G using the CMC. But while the CFC is often deployed as a simplifying assumption in causal inference,[6] it is universally acknowledged to fall prey to counterexamples. Indeed, there are some probability distributions that cannot be paired with any causal graph to satisfy both the CMC and the CFC.[7] Thus it would seem that we can't rely on the CFC in our justification of any fully general account of causal direction in causal Bayes nets.

Of particular interest here is a class of distributions defined over three variables $\{V_1, V_2, V_3\}$ in which the only independencies that obtain are $V_1 \perp\!\!\!\perp V_3$ and $V_1 \perp\!\!\!\perp V_3 | V_2$.[8] Though there is no causal graph that can be paired with such a distribution to satisfy both the CMC and the CFC,[9] it is easy to imagine multiple causal scenarios that give rise to such distributions—even when we stipulate that V_1 is exogenous in order to mirror the assumptions of our inference problem in which E_x is known to be exogenous. To make things concrete, let us first identify a specific probability distribution according to which these independencies obtain, depicted in table 1.

Note that this distribution satisfies the following properties:

- $\neg(V_1 \perp\!\!\!\perp V_2)$, since $P(V_2 = 1 | V_1 = 0) = \frac{10}{33} \neq \frac{20}{33} = P(V_2 = 1 | V_1 = 1)$.

[6] This assumption is often defended on the grounds that it will seldom lead us astray, but there are some domains in which violations of the CFC are ubiquitous—e.g., when the values of the variables at play enter into deterministic relations.

[7] See Zhang's (2013) discussion of "detectable failures of faithfulness."

[8] These distributions are discussed at length by Zhang (2013), albeit in the context of constraint-based causal search.

[9] See Zhang (2013), pp. 430–431.

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

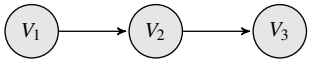


Figure 1. An intransitive chain.

- $\neg(V_2 \perp\!\!\!\perp V_3)$, since $P(V_3 = 1|V_2 = 2) = \frac{9}{10} \neq P(V_3 = 1) = \frac{56}{66}$.
- $V_1 \perp\!\!\!\perp V_3$, since $P(V_3 = 1) = \frac{28}{33} = P(V_3 = 1|V_1 = 0)$.
- $V_1 \perp\!\!\!\perp V_3|V_2$, since

$$P(V_3 = 1|V_1 = 0, V_2 = 0) = \tfrac{1}{3} = P(V_3 = 1|V_1 = 1, V_2 = 0),$$

$$P(V_3 = 1|V_1 = 0, V_2 = 1) = \tfrac{9}{10} = P(V_3 = 1|V_1 = 1, V_2 = 1),$$

$$P(V_3 = 1|V_1 = 0, V_2 = 2) = \tfrac{9}{10} = P(V_3 = 1|V_1 = 1, V_2 = 2).$$

The interesting thing about this distribution is that it represents V_1 as being correlated with V_2, V_2 as being correlated with V_3, and V_1 being independent of V_3 *both* unconditionally and conditional on any value of V_2. To see how this happens, note that while V_1 and V_3 are both binary variables, V_2 is ternary. Conditioning on $V_1 = 0$ makes no difference to the probability that $V_2 = 0$, but it does make a difference to the probabilities of $V_2 = 1$ and $V_2 = 2$. However, the only way in which V_2 is relevant to V_3 is that V_3's probability changes depending on whether $V_2 = 0$—i.e., if $V_2 \neq 0$, it makes no difference to V_3 whether $V_2 = 1$ or $V_2 = 2$.

Because this distribution has these properties, it is possibly realized by a slight variant of McDermott's (1995) famous "dog bite" counterexample to the transitivity of causation.[10] In this scenario, a terrorist is contemplating pressing a button that will probably detonate a bomb, which will probably blow up a football stadium, but there is a dog present who threatens to bite the terrorist's right hand. As things turn out, the probability that the terrorist detonates the bomb is not affected by whether the dog bites, but the probability that the terrorist uses her right or left hand (in the event that she pushes the button) is affected. If we allow V_1 to represent whether the dog bites, V_2 to represent whether the button is pressed, and if so with which hand, and V_3 to represent whether the explosion occurs, then our example distribution provides a natural model of this situation.[11] V_1 is correlated with V_2 because the dog biting makes a difference to which hand the terrorist uses (if they push). V_2 is correlated with V_3 because pushing (with either hand) obviously increases the probability of an explosion. And V_3 is probabilistically independent of V_1 both unconditionally and conditional on any value of V_2 because V_1 can only make a difference to which hand is used, which is irrelevant to whether there's an explosion (which depends only on whether the button is pushed at all). Figure 1 represents the direct causal relationships that intuitively obtain in the example.

[10] The only difference is that there is no local determinism in the model. Zhang and Spirtes (2008) likewise discuss a version of McDermott's case in which there is no local determinism.

[11] Here, V_1 is 0 when the dog does not bite and 1 when the dog does bite. V_2 is 0 when the button is not pushed, 1 when the button is pushed with the right hand, and 2 when the button is pushed with the left hand. V_3 is 0 when there is no explosion and 1 when there is an explosion.

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

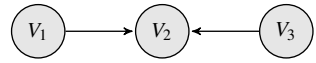


Figure 2. An unidentifiable collider.

Importantly, this probability distribution can likewise be realized by a somewhat similar scenario where the arrow between V_2 and V_3 goes in the opposite direction. Let V_1 again represent whether the dog bites, V_2 again represent whether the button is pressed, and if so with which hand, but allow V_3 to now represent whether the terrorist receives orders from her superior to push the button. This scenario likewise is intuitively consistent with the probability distribution under consideration since nothing seems to block us from treating the explosion from the first example as probabilistically indistinguishable from receiving orders in the second example—i.e., it could be that when the terrorist pushes the button, it is probably because she received orders to do so, but also that the probability that she received orders is not at all further affected by whether she pushes with her right hand or her left hand.[12] But in this case, the intuitive causal graph over the very same distribution is now as shown in Figure 2.[13]

At this stage, it's useful to think through the way in which both of these candidate causal structures generate violations of the CFC. The structure in Figure 1 violates the CFC with respect to our distribution because it does not entail (via the CMC) that $V_3 \perp\!\!\!\perp V_1$, even though V_3 is in fact unconditionally independent of V_1 in the distribution. The structure in Figure 2 violates the CFC with respect to our distribution because it does not entail (via the CMC) that $V_1 \perp\!\!\!\perp V_3 | V_2$, even though V_1 is in fact conditionally independent of V_3 given V_2 in the distribution. In the parlance of graphical causal models, this is a case where conditioning on a collider fails to induce a correlation between two unconditionally independent causes. The examples here illustrate that intransitive causal chains bear the same probabilistic signature as cases where conditioning on a collider fails to induce a correlation, and that both of these phenomena yield violations of the CFC.[14] Now, since both of the candidate causal structures are ruled out by the CFC, one might hope to appeal to a less demanding condition as we consider these cases. Enter the significantly weaker and arguably axiomatic Causal Minimality Condition (CMIN).

The CMIN says that a causal graph G over $\mathbf{V}$ is compatible with a probability distribution $\mathbb{P}$ over $\mathbf{V}$ exactly when there exists no proper sub-graph of G that satisfies the CMC with $\mathbb{P}$, where a graph qualifies as a proper sub-graph of G if and only if (i) it excludes some arrow(s) in G, and (ii) nothing else is changed (such that all remaining arrows are oriented in the same direction as in G). Intuitively, the CMIN requires that

[12] Clearly, both the orders and the explosion are strongly correlated with whether the button is pushed, but are independent of whether the dog bites, both unconditionally and conditional on any value of V_2.

[13] We will see in what follows that this kind of underdetermination is important in the present context because both Figure 1 and Figure 2 are consistent with the assumption that V_1 is exogenous, and thereby reflect scenarios where V_1 effectively plays the role of an exogenous source of variation on V_2.

[14] We are not aware of the probabilistic indistinguishability of intransitive causal chains and "unidentifiable colliders" of this sort being recognized anywhere in the extant literature. It is, however, commonly acknowledged that the information contained witihin a causal graph goes beyond its probabilistic implications (see, e.g., Eva, Stern, and Hartmann (2019)).

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

a causal graph should include *just enough* arrows to guarantee satisfaction of the CMC, and should not include any other arrows that are not necessary for that purpose. It's not hard to see that the structures in Figures 1 and 2 both satisfy the CMIN with respect to the given distribution. Since both V_1 and V_3 are pairwise unconditionally correlated with V_2, deleting either of the arrows in either structure will yield a violation of the CMC by entailing the unconditional independence of V_2 with one of the other two variables. Ergo, the CMIN manages to capture the sentiment that both of these causal structures are compatible with the given the probability distribution, since both satisfy both the CMC and the CMIN when paired with the given distribution.[15]

The upshot of this is that we can't identify the direction of the arrow between V_2 and V_3 simply by consulting the probability distribution over V_1, V_2, and V_3 when V_1 is known to be exogenous. More generally, when the only conditional independencies that obtain are $V_1 \perp\!\!\!\perp V_3$ and $V_1 \perp\!\!\!\perp V_3 | V_2$, then the set of causal graphs that can be paired with the distribution to satisfy the CMC and the CMIN are $V_1 \rightarrow V_2, \rightarrow V_3$, $V_1 \leftarrow V_2, \rightarrow V_3$, $V_1 \leftarrow V_2, \leftarrow V_3$, and $V_1 \rightarrow V_2, \leftarrow V_3$.[16] Thus, while the CMC and the CMIN combine to tell us something about how many arrows there must be (since any graph with three arrows would fail to qualify as minimal and any graph with one or fewer arrows would fail to qualify as Markovian), they jointly tell us nothing about the direction of the arrows that there are.

Of course, it's old news that we cannot derive causal order from merely probabilistic information since, e.g., $V_1 \rightarrow V_2, \rightarrow V_3$, $V_1 \leftarrow V_2, \rightarrow V_3$, and $V_1 \leftarrow V_2, \leftarrow V_3$ form a "Markov equivalence class" (meaning that they entail the same independencies via the CMC) and are therefore indistinguishable even given the CFC. But the kind of underdetermination at play here is of an especially vexing variety since causal graphs that are *not* Markov equivalent end up being indistinguishable, even when they contain a fixed number of arrows.[17] In the present context, this is pressing because the assumption that the source of variation V_1 is exogenous is

[15] We have chosen to focus on the CMIN here because we believe that the CMIN has a genuine claim to axiomaticity, but it's worth noting that both the Frugality Condition and the Pearl Minimality Condition likewise sanction Figures 1 and 2 as compatible with the probability distribution under consideration. The Frugality Condition says that the true causal graph must be among the set of Markovian graphs that minimize the number of arrows included therein. Here, it is easy to see that there is no Markovian graph with strictly fewer arrows than Figures 1 or 2. And since the Pearl Minimality Condition is strictly weaker than the Frugality Condition, it follows that both graphs are likewise sanctioned as admissible by Pearl Minimality. This effectively means that there is no way to rule out one of these causal graphs by relaxing the CFC since every relevant weakening of the CFC that is proposed in the literature sanctions both causal graphs as admissible. See Forster et al. (2018) for extensive discussion of the relationship between these conditions.

[16] All four graphs are also admissible given the distribution if we replace the CMIN with the Frugality Condition or the Pearl Minimality Condition.

[17] The structures in Figures 1 and 2 are *not* Markov equivalent because, e.g., the structure in Figure 1 entails that $V_1 \perp\!\!\!\perp V_3 | V_2$, while the structure in Figure 2 does not. Philosophers sometimes discuss cases where two causal graphs that are not Markov equivalent are compatible with the same probability distributions, where the phenomenon is due to the possibility that multiple paths probabilistically cancel each other out in one graph to mimic the conditional independencies that are implied (via the CMC) by another graph in which there are fewer arrows. The kind of underdetermination at play in our examples is importantly different from this kind, since each of the admissible graphs contains the same number of arrows. As Forster et al. (2018) would describe the situation, this means that the kind of

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

sufficient to rule out all but $V_1 \rightarrow V_2, \rightarrow V_3$ when V_1 is probabilistically independent of V_3, given the CFC. But when we relax the CFC so that we can model our examples (and instead assume the CMIN), both $V_1 \rightarrow V_2, \rightarrow V_3$ and $V_1 \rightarrow V_2, \leftarrow V_3$ are compatible with V_1's exogeneity and probabilistic independence with V_3. Thus, when the setting is general enough to incorporate examples like these, we cannot always infer causal direction from exogenous variation.

3. A blocked escape route

Might you always be able to solve the problem posed by these cases by additionally looking at what happens when there is an exogenous source of variation of Y (or V_3)? No. Here's why. Recall the example where V_1, V_2, and V_3 refer to the dog bite / button pushing / explosion variables. Now fine-grain the values of V_3 further so that instead of simply distinguishing between the explode / don't explode possibilities, V_3 now distinguishes between *three* possibilities, namely (i) stadium explosion, (ii) no stadium explosion and tomorrow's football match happens as scheduled, and (iii) no stadium explosion and tomorrow's football match is canceled. Now, let V_4 be an exogenous cause of V_3 that denotes whether there are severe thunderstorms tomorrow. Here, V_4 is pairwise correlated with V_3 and V_3 is pairwise correlated with V_2, but we can imagine that V_4 is both conditionally and unconditionally uncorrelated with V_2 since tomorrow's weather isn't predictive of how and whether the button gets pushed, no matter whether we condition on any of the values of V_3.[18]

This means that if our problem is to infer the direction of causation between V_2 and V_3, then we are in the same position as before. For when we include an exogenous source of variation of V_3 (namely, V_4) in the causal graph, it turns out that the CMC and the CMIN sanction both $V_2 \rightarrow V_3 \rightarrow V_4$ and $V_2 \rightarrow V_3 \leftarrow V_4$ as compatible with the distribution at hand.

The upshot of this is that there are cases where looking at exogenous sources of variation of both X and Y is not sufficient for unveiling the direction of the causal relationship between X and Y. To see this, it is helpful to consider all four of the variables E_X, X, Y, and E_Y as a single variable set (rather than just considering the subsets $\{E_X, X, Y\}$ and $\{E_Y, X, Y\}$ in isolation. In our running example, this means looking at a distribution that is defined over V_1, V_2, V_3, and V_4 (where V_3 is ternary). In this distribution, (i) V_2 is unconditionally pairwise correlated with both V_1 and V_3, (ii) V_3 is unconditionally pairwise correlated with both V_2 and V_4, (iii) V_1 is independent of V_4 both unconditionally and conditional on both V_2 and V_3, (iv) V_1 is independent of V_3 both unconditionally and conditional on V_2, and (v) V_4 is independent of V_2 both unconditionally and conditional on V_3.[19] Now consider the possible causal structures over this four variable set, as depicted in Figures 3 and 4.

underdetermination at play is not made possible by one graph implying fewer basic independencies than another, and thereby being able to accommodate more probability distributions than the other.

[18] Plausibly, the weather doesn't tell you anything about whether (and how) the button gets pushed, and it doesn't seem that learning whether there's an explosion (and if there's not, whether the game goes ahead) changes that.

[19] Phrased in terms of our running example, V_1 is the binary dog bite variable, V_2 is the ternary button pushing variable, V_3 is the ternary explosion / football game variable, and V_4 is the binary weather variable. Of course, in the story we gave, we know that Figure 3 depicts the true causal structure, but the

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

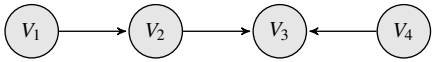


Figure 3. Four variable structure 1.

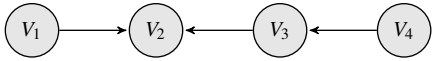


Figure 4. Four variable structure 2.

Given the probability distribution we just described, both of these causal structures satisfy both the CMC and the CMIN since they don't entail any false independencies via the CMC, but deleting any arrows would force them to do so. Now, in the case where we are considering the variable set $\{E_Y, E_Y, X, Y\}$ with the hope of discerning the direction of causation between X and Y, we need to distinguish between the structures $E_X \rightarrow X \rightarrow Y \leftarrow E_Y$ and $E_X \rightarrow X \leftarrow Y \leftarrow E_Y$. But if we're in a situation like the one just described, we won't be able to distinguish between these two structures using the CMC and the CMIN alone (since doing so requires being able to discern between the structures depicted in Figures 3 and 4). Put differently, when our problem is to infer the direction of non-spurious causation between V_2 and V_3, there is no guarantee that we'll be able to solve this problem simply by embedding V_2 and V_3 in a distribution that includes exogenous sources of variation for both, since this leaves open the possibility that the distribution at hand will include the conditional independencies that we've described over V_1, V_2, V_3, and V_4—in which case both Figures 3 and 4 will be admissible. So overall, there is simply no guarantee that considering an exogenous cause of Y as well as an exogenous cause of X will help us to determine the direction of causation between X and Y.

4. Possible escape routes

Does this spell doom for this approach to understanding causal direction in causal Bayes nets?

Maybe. If so, then perhaps we should revisit the temporal order strategy. We can show that if causes always temporally precede their effects, then if V_1, V_2, and V_3 are temporally ordered, there is a uniquely admissible causal graph given the CMC and the CMIN in the cases that concern us here.[20] For instance, if V_2 temporally precedes V_3 and we assume that (i) V_2 and V_3 are non-spuriously correlated, (ii) V_2 temporally precedes V_3, and (iii) causes always temporally precede their effects, it follows that $V_1 \rightarrow V_2 \rightarrow V_3$ is admissible but $V_1 \rightarrow V_2 \leftarrow V_3$ is not—i.e., we can infer the direction of causation between V_2 and V_3. But when we opt for this route, we abandon the project of understanding the asymmetry of causation in terms of any probabilistic (in)dependencies since the inferential work is ultimately accomplished by the

point is that one could not distinguish between Figures 3 and 4 based on probabilities alone, even when one knows that V_1 and V_4 are exogenous causes of V_2 and V_3, respectively.

[20] More generally, we can show that this inferential strategy will always yield a unique graph, provided that the probability distribution over $\mathbf{V}$ assigns positive probability to every combination of values of the variables in $\mathbf{V}$. See Hitchcock (2018), Pearl (1988), and Stern (2021) for extensive discussion of this result and its relevance to causal inference.

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

temporal ordering, not any probabilistic (in)dependence that is secured by the CMC or the CMIN. Moreover, this would seem to involve maintaining that causes temporally precede their effects as a matter of conceptual necessity. As we mentioned at the outset, it is controversial whether this is true. But even if we grant this conceptual claim to the defender of temporal precedence, it doesn't resolve the current issue in every realistic inference problem since we are often confronted with variable sets whose temporal ordering is unknown to us. Thus, even if relying on temporal orderings is a good strategy whenever said orderings are available, the question of what to do when they are unavailable remains.

There may also be other ways to solve the problem that do not invoke temporal order. One possibility is that we can avoid the problematic examples discussed here by restricting the domain of the causal relation to binary variables. This is not frequently discussed in the philosophical literature (Spohn 2001; 2012 are exceptions), but it's easy to see how this question arises against the backdrop of a contrastivist account of causation.[21] Toward this end, Schaffer (2010) has argued that when we treat variables as the *relata* of the causal relation, we take a step toward agreeing with the contrastivist that the causal relation is at least a four-place relation, since saying that X causes Y may just be another way of saying that one value of X rather than another causes one value of Y rather than another. But if variables with more than two values can be causal *relata*, then as Schaffer (2010, fn. 28) notes, the contrastivist framework must be extended to allow for sets of causal contrasts and sets of effectual contrasts. One way to defend using exogenous variation as a general strategy for inferring causal direction may involve arguing that any such extension would be unprincipled.

Finally, a third possible escape route that one could explore would be to identify a special sub-class of exogenous causes which have properties that make it impossible for the kinds of examples we've discussed here to arise. In particular, one could try to show that whenever E_X is a certain kind of exogenous cause of X and X is non-spuriously correlated with Y, it will be impossible for E_X, X, and Y to stand in the kind of probabilistic relationship that V_1, V_2, and V_3 stand in in our earlier examples. A natural starting place here would be to consider idealized exogenous interventions that deterministically set the values of their direct effects. It may turn out that such interventions do indeed avoid these kinds of problematic examples.[22] But even if that is the case, the question remains of whether there exists any less idealized subset of

[21] Spohn (2001; 2012) maintains that causes temporally precede their effects, and that admissible variables are both binary and temporally ordered, but it seems possible that either one of these restrictions on their own could be sufficient to avoid the issues described here.

[22] The possibility of such a result is likely to depend on exactly how one characterizes the notion of an ideal intervention. If we follow Pearl (2009) in characterizing interventions in terms of the do-operator, it is unclear how we can even begin to work towards such a result, since there are no probabilities assigned to intervention variables in this setup. There is more hope if we instead follow Spirtes, Glymour, and Scheines (2000) in characterizing interventions as variables that we condition upon in order to arrive at the distributions that result from intervening. But here, too, it isn't obvious how to proceed since this involves assigning zero unconditional probability to every value of the intervention variable that corresponds to intervening on the target variable and maximal unconditional probability to the value that corresponds to not intervening (so as to yield the unmanipulated observational distribution). This means that any result of this sort would be proved in a non-classical setting wherein the ratio formula definition of conditional probability doesn't hold.

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

exogenous causes that allow one to avoid the problem. That is, since it is acknowledged by many that there are examples where no ideal "hard" intervention is adequate to the inference problem at issue,[23] it would be good to know exactly what kinds of "soft" interventions are suitable for inferring the direction of causation in cases like the ones described in this paper.

5. Conclusion

In summary, we've pointed out:

(1) that there can exist Markov inequivalent graphs with the same number of arrows that cannot be distinguished by the combination of the CMC and any extant weakening of CFC;
(2) that this causes fundamental problems for inferring the direction of causation in causal Bayes nets from probabilistic (in)dependence with exogenous causes;[24] and
(3) that one might hope to resolve the problem by either (i) restricting one's attention exclusively to binary variables, (ii) relying on temporal orderings, or (iii) identifying special sub-classes of exogenous causes that preclude the possibility of the examples discussed here.

Acknowledgments. For helpful discussion and comments, we are grateful to Malcolm Forster, Olav Vassend, Jiji Zhang, and the audience at the 2024 meeting of the Society for the Philosophy of Causation.

Funding and declarations. None to declare.

References

Eberhardt, Frederick. 2014. "Direct Causes and the Trouble with Soft Interventions." *Erkenntnis* 79 (4): 1–23. doi: 10.1007/s10670-013-9552-2.

Eva, Benjamin, Reuben Stern, and Stephan Hartmann. 2019. "The Similarity of Causal Structure." *Philosophy of Science* 86 (15):821–35. doi: 10.1086/705566.

Hausman, Daniel M., and James Woodward. 1999. "Independence, Invariance and the Causal Markov Condition." *British Journal for the Philosophy of Science* 50 (4):521–83. doi: 10.1093/bjps/50.4.521.

Hitchcock, Christopher. 2018. "Causal Models." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford, CA: Stanford University Press. https://plato.stanford.edu/entries/causal-models/.

Korb, Kevin B., Lucas R. Hope, Ann E. Nicholson, and Karl Axnick. 2004. "Varieties of Causal Intervention." In *PRICAI 2004: Trends in Artificial Intelligence*, edited by Chenqi Zhang, Hans W. Guesgen, and Wai-Kiang Yeap, 322–331. Berlin: Springer. doi: 10.1007/978-3-540-28633-2_35.

Forster. Malcolm, Garvesh Raskutti, Reuben Stern, and Naftali Weinberger. 2018. "Frugal Inference of Causal Relations." *British Journal for the Philosophy of Science* 69 (3):821–48. doi: 10.1093/bjps/axw033.

McDermott, Michael. 1995. "Redundant Causation." *British Journal for the Philosophy of Science* 40: 523–544. https://doi.org/10.1093/bjps/46.4.523.

Pearl, Judea. 1988. *Probabilsitic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Burlington, MA: Morgan Kaufmann.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*, 2nd edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511803161.

[23] See, e.g., Eberhardt (2014) and Korb et al. (2004).

[24] Our observations towards this end are clearly relevant to Woodward's recent (2022) paper on inferring causal direction from observational data, as well as to the rich body of work in machine learning with which Woodward engages. See, e.g., Zhang, Zhang, and Schölkopf (2015). We have refrained from engaging with this work because it lies beyond the purview of this paper, but it is something that we hope to address in future work.

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press

Price, Huw. 1996. *Time's Arrow and Archimedes' Point*. Oxford: Oxford University Press.
Schaffer, Jonathan. 2010. "Contrastive Causation in the Law." *Legal Theory* 16 (4):259–97. doi: 10.1017/S1352325210000224.
Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction and Search*. Cambridge, MA: MIT Press. doi: https://doi.org/10.7551/mitpress/1754.001.0001.
Spohn, Wolfgang. 2001. "Bayesian Nets Are All There Is To Causal Dependence." In *Stochastic Causality*, edited by Maria Carla Galavotti, Patrick Suppes, and Domenico Constantini, 157–72. Stanford, CA: CSLI Publications.
Spohn, Wolfgang. 2012. *The Laws of Belief: Ranking Theory and its Philosophical Applications*. Oxford: Oxford University Press. doi: https://doi.org/10.1093/acprof:oso/9780199697502.001.0001.
Stern, Reuben. 2021. "Causal Concepts and Temporal Ordering." *Synthese* 198:6505–27. doi: https://doi.org/10.1007/s11229-019-02235-4.
Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press. doi: 10.1093/0195155270.001.0001.
Woodward, James. 2022. "Flagpoles Anyone? Causal and Explanatory Asymmetries." *Theoria* 37 (1):7–52.
Zhang, Jiji. 2013. "A Comparison of Three Occam's Razors for Markovian Causal Models." *British Journal for the Philosophy of Science* 64 (2):423–48. doi: https://doi.org/10.1093/bjps/axs005.
Zhang, Jiji, and Peter Spirtes. 2008. "Detection of Unfaithfulness and Robust Causal Inference." *Minds and Machines* 18: 239–71. doi: https://doi.org/10.1007/s11023-008-9096-4.
Zhang, Kun, Jiji Zhang, and Bernhard Schölkopf. 2015. "DIstinguishing Cause from Effect Based On Exogeneity." In *Proceedings of the Fifteenth Conference on Theoretical Aspects of Rationality and Knowledge* edited by R. Ramanujam, 261–271. doi: https://doi.org/10.48550/arXiv.1504.05651.

Cite this article: Stern, Reuben and Benjamin Eva. 2026. "Causal Direction in Causal Bayes Nets." *Philosophy of Science* 93 (1):170–180. https://doi.org/10.1017/psa.2025.10102

https://doi.org/10.1017/psa.2025.10102 Published online by Cambridge University Press