Experimental Indistinguishability of Causal Structures

Frederick Eberhardt*

Using a variety of different results from the literature, I show how causal discovery with experiments is limited unless substantive assumptions about the underlying causal structure are made. These results undermine the view that experiments, such as randomized controlled trials, can independently provide a gold standard for causal discovery. Moreover, I present a concrete example in which causal underdetermination persists despite exhaustive experimentation and argue that such cases undermine the appeal of an interventionist account of causation as its dependence on other assumptions is not spelled out.

1. Introduction. Causal search algorithms based on the causal Bayes net representation (Pearl 2000; Spirtes, Glymour, and Scheines 2000) have primarily focused on the identification of causal structure using passive observational data. The algorithms build on assumptions that connect the causal structure represented by a directed (acyclic) graph among a set of vertices with the probability distribution of the data generated by the causal structure. Two of the most common such bridge principles are the causal Markov assumption and the causal faithfulness assumption. The *causal Markov assumption* states that each causal variable is probabilistically independent of its (graphical) nondescendents given its (graphical) parents. Causal Markov enables the inference from a probabilistic dependence between two variables to a causal connection and from a causal separation to a statistical independence. The precise nature of such causal separation and connection relations is fully characterized by the notion of d-separation (Geiger, Verma, and Pearl 1990; Spirtes et al. 2000, 3.7.1). The *causal faithfulness assumption* can be seen as the converse to the Markov assumption. It states that all and only the independence relations true in the probability distribution over the set of variables are a consequence of the Markov condition. Thus, faithfulness

*To contact the author, please write to: California Institute of Technology; e-mail: fde@caltech.edu.

Philosophy of Science, 80 (December 2013) pp. 684–696. 0031-8248/2013/8005-0013$10.00
Copyright 2013 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/673865 Published online by Cambridge University Press

permits the inference from probabilistic independence to causal separation and from causal connection to probabilistic dependence. Together causal Markov and faithfulness provide the basis for causal search algorithms based on passive observational data. For the simplest case they are combined with the assumptions that the causal structure is acyclic and that the measured variables are causally sufficient, that is, that there are no unmeasured common causes of the measured variables. For example, given three variables x, y, and z, if we find that the only (conditional or unconditional) independence relation that holds among the three variables is that x is independent of z given y, then causal Markov and faithfulness allow us to infer that the true causal structure is one of those presented in figure 1.

Causal Markov and faithfulness do not determine which of the three causal structures is true, but this underdetermination is well understood for causal structures in general. It is characterized by the so-called Markov equivalence classes of causal structures. These equivalence classes consist of sets of causal structures (graphs) that have the same independence and dependence relations among the variables. The three structures in figure 1 are one such equivalence class. To identify the true causal structure uniquely there are two options: one can make stronger assumptions about the underlying causal model or one can run experiments. Here I will first focus on the latter to then show that one cannot really do without the former.

I will take an experiment to consist of an intervention on a subset of the variables under consideration. While there are a variety of different types of interventions, I will focus here on experiments involving so-called surgical interventions (Pearl 2000). In a *surgical intervention* the intervention completely determines the probability distribution of the intervened variable and thereby makes it independent of its normal causes. Such an intervention is achieved (at least in principle) by a randomized controlled trial: whether or not a particular treatment is administered is determined entirely by the randomizing device and not by any other factors. In a causal Bayes net a surgical intervention breaks the arrows into the intervened variable, while leaving the remaining causal structure intact. It is possible to perform an experiment that surgically intervenes on several variables simultaneously and independently. In that case, of course, all information about the causal relation among intervened variables is lost.

For the three Markov equivalent structures in figure 1, a single-intervention experiment intervening only on y would distinguish the three causal structures: it would make x independent of y if the first structure is true, but not

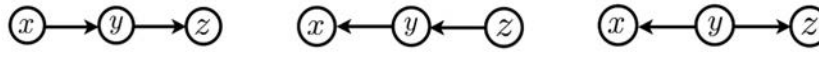


Figure 1.

https://doi.org/10.1086/673865 Published online by Cambridge University Press

for the second and third. And it would make y independent of z if the second structure is true, but not for the first and the third. Together these two considerations show that such an experiment on y would resolve the underdetermination of this Markov equivalence class completely.

Ever since Ronald A. Fisher's work in the 1930s, experiments have come to be seen as the gold standard for causal discovery (Fisher 1935). This view suggests that if one can perform experiments, then causal discovery is (theoretically) trivial. The recent rise of the interventionist account of causation in philosophy appears to endorse this view, since it holds that just *what it is* to stand in a causal relation, is the possibility of performing the appropriate kind of experiment (Woodward 2003).

2. Underdetermination despite Experiments. First the hopeful news: Eberhardt, Glymour, and Scheines (2005) showed that one can generalize the strategy used to identify the true causal structure in figure 1 to arbitrary causal structures over N variables: assuming that causal Markov, faithfulness, and causal sufficiency hold, and that the causal structure is acyclic, one can uniquely identify the true causal structure among a set of variables given a set of single-intervention experiments. Generally such a procedure will require several experiments intervening on different variables, but a sequence of experiments that guarantees success can be specified.

Similar results can be obtained without experiments but by instead strengthening the assumptions one makes about the underlying causal structure. Shimizu et al. (2006) show that if causal sufficiency holds, the causal relations are linear, and the error distributions on the variables are non-Gaussian, then the causal structures can also be uniquely identified. A set of causal variables is related linearly when the value of each variable is determined by a linear function of the values of its parents plus an error term. Each error variable has a disturbance distribution, and as long as these distributions are not Gaussian (and not degenerate), then the same identifiability of causal structure is guaranteed as would be obtained by not making the assumptions about the causal relations, but instead running a set of single-intervention experiments.

In either case, whether by strengthening assumptions or using experiments, the results rely on the assumption of causal sufficiency—that there are no unmeasured common causes. In many discovery contexts it is implausible that such an assumption is appropriate. Moreover, part of the rationale for randomized controlled trials in the first place was that a randomization makes the intervened variable independent of its normal causes, whether those causes were measured or not. Thus, if there is an unmeasured common cause u—a confounder—of x and z, then randomizing x would break the (spurious) correlation observed between x and z that is due to

https://doi.org/10.1086/673865 Published online by Cambridge University Press

the confounder u. However, without the assumption of causal sufficiency, underdetermination returns despite the possibility of experiments.

In figure 2, x, y, and z are observed (and can be subject to intervention), while u and v are unobserved. If only causal Markov, faithfulness, and acyclicity are assumed, the two causal structures in figure 2 cannot be distinguished by any set of experiments that intervene on only one variable in each experiment (or by a passive observation). Since u and v are not observed, no variable is (conditionally) independent of any other variable under passive observation. The same is true when x is subject to an intervention, even though the surgical intervention would break the influence of u on x: x is not independent of z conditional on y, since conditioning on y induces a dependence via v (conditioning on a common effect makes the parents dependent). In an experiment intervening on y only, x and y are independent, but x and z remain dependent for both causal structures (because of u in structure 2 and because of u and the direct effect $x \rightarrow z$ in structure 1). In an experiment intervening on z, the edge $x \rightarrow z$ that distinguishes the two causal structures is broken, so both structures inevitably have the same independence and dependence relations. The problem is that no set of single-intervention experiments is sufficient to isolate the $x \rightarrow z$ edge in structure 1, and so the underdetermination remains.

This underdetermination can, of course, be resolved: if one could intervene on x and y simultaneously, then x will be independent of z if the second structure is true, but dependent if the first is true. So, assuming only causal Markov, faithfulness, and acyclicity, the two causal structures are experimentally indistinguishable for single-intervention experiments but distinguishable for double-intervention experiments.

How does this generalize to arbitrary causal structures? The resolution of the underdetermination of the causal structures in figure 2 depended on an experiment that intervened on all but one variable simultaneously. This is true in general: assuming causal Markov, faithfulness, and acyclicity, but not

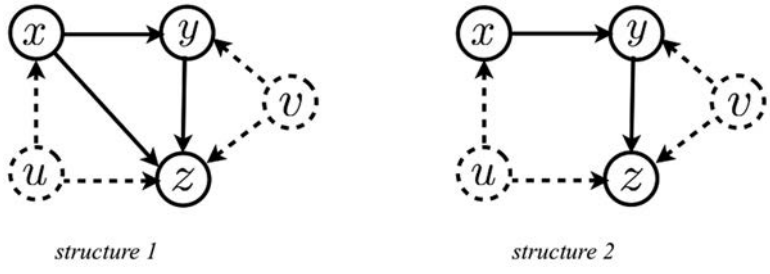


Figure 2.

https://doi.org/10.1086/673865 Published online by Cambridge University Press

causal sufficiency, there exist at least two causal structures over N variables that are indistinguishable on the basis of the independence and dependence structure for all experiments that intervene on at most $N-2$ variables, where N is the number of observed variables. That is, at least one experiment intervening on all but one variable is necessary to uniquely identify the true causal structure. In fact, the situation is worse, because a whole set of experiments, each intervening on at least $N-i$ variables, for each integer i in $0 < i < n$, is in the worst case necessary to ensure that the underdetermination is resolved (see app. A for a proof). So, even when multiple simultaneous interventions are possible, a large number of experiments each intervening on a large number of variables simultaneously are necessary to resolve the underdetermination.

Again, one need not pursue this route. One could instead strengthen the search space assumptions. Part of why single-intervention experiments were not sufficient to resolve the underdetermination of the causal structures in figure 2 is that independence tests are a general but crude tool of analysis. If one could separate the causal effect of the $x \to y \to z$ pathway from the direct causal effect of $x \to z$ in the structures in figure 2, then the two causal structures could be distinguished. For linear causal relations this is possible, since so-called trek rules specify that the correlation between two variables in a linear model is given by the sum-product of the correlations along the (active) treks that connect the variables. Suppose that the linear coefficient of the $x \to y$ edge is a, of the $y \to z$ edge is b and of the $x \to z$ edge is c in figure 2. If the second structure is true, then in an experiment that intervenes on x, we have $\mathrm{cor}(x, z) = ab$, while if the first structure is true, then $\mathrm{cor}(x, z) = ab + c$ in the same experiment. We can measure the correlations and compare the result to the predictions: in an experiment that intervenes on y, we can determine b by measuring $\mathrm{cor}(y, z)$. In an experiment intervening on x, we can determine a by measuring $\mathrm{cor}(x, y)$, and we can measure $\mathrm{cor}(x, z)$. If $\mathrm{cor}(x, z) = \mathrm{cor}(x, y)\mathrm{cor}(y, z) = ab$, then the second structure is true, while if the first structure is true, then $\mathrm{cor}(x, z) \neq \mathrm{cor}(x, y)\mathrm{cor}(y, z)$, and we can determine $c = \mathrm{cor}(x, z) - \mathrm{cor}(x, y)\mathrm{cor}(y, z)$. Thus, on the basis of single-intervention experiments alone we are able to resolve the underdertermination. But we had to assume linearity.

Eberhardt, Hoyer, and Scheines (2010) show that this approach generalizes: if the causal model is linear (with any nondegenerate distribution on the error terms), but causal sufficiency does not hold, then there is a set of single-intervention experiments that can be used to uniquely identify the true causal structure among a set of variables. This result holds even when the assumptions of acyclicity and faithfulness are dropped. It shows just how powerful the assumption of linearity is. Linearity is sufficient to achieve identifiability even for single-intervention experiments, but it is known not

https://doi.org/10.1086/673865 Published online by Cambridge University Press

to be necessary. Hyttinen, Eberhardt, and Hoyer (2011) have shown that similar results can be achieved for particular types of discrete models—so-called noisy-or models. It is currently not known what type of parametric assumption is necessary to avoid single-intervention experimental indistinguishability.

However, there is a weaker result: appendix B contains two discrete (but faithful) parameterizations, one for each of the causal structures in figure 2 (adapted from Hyttinen et al. 2011). I refer to the parameterized model corresponding to the first structure as PM1 and that for the second structure as PM2. As can be verified from appendix B, PM1 and PM2 have identical passive observational distributions, identical manipulated distributions for an experiment intervening only on x, an experiment intervening only on y, and (unsurprisingly) for an experiment intervening only on z. That is, the two parameterized models are not only indistinguishable on the basis of independence and dependence tests for any single-intervention experiment or passive observation. They are indistinguishable in principle, that is, for any statistical tool, given only single-intervention experiments (and passive observation), because those (experimental) distributions are identical for the two models. This underdetermination exists despite the fact that all (experimental) distributions are faithful to the underlying causal structure. The models are, however, distinguishable in a double-intervention experiment intervening on x and y simultaneously. Only for such an experiment do the experimental distributions differ so that the presence of the $x \rightarrow z$ edge in PM1 is in principle detectable. I do not know, but conjecture, that this in-principle-underdetermination (rather than just the underdetermination based on the [in-]dependence structure, as shown in app. A) can be generalized to arbitrary numbers of variables and will hold for any set of experiments that at most intervene on $N - 2$ variables.

The example shows that in order to identify the causal structure by single-intervention experiments some additional parametric assumption beyond Markov, faithfulness, and acyclicity is necessary. Alternatively, without additional assumptions, causal discovery requires a large set of very demanding experiments, each intervening on a large number of variables simultaneously. For many fields of study it is not clear that such experiments are feasible, let alone affordable or ethically acceptable. It is unclear how common cases like PM1 and PM2 are. It is possible that in practice such cases are quite rare. When the assumption of faithfulness was subject to philosophical scrutiny, one argument in its defense was that a failure of faithfulness was for certain types of parameterizations a measure-zero event (Spirtes et al. 2000, theorem 3.2). While this defense of faithfulness has not received much philosophical sympathy, such assessments of the likelihood of trouble are of interest when one is willing or forced to make the antecedent parametric as-

https://doi.org/10.1086/673865 Published online by Cambridge University Press

sumptions anyway. The example here does not involve a violation of faithfulness, but a similar analysis of the likelihood of underdetermination despite experimentation is possible.

PM1 and PM2 cast a rather dark shadow on the hopes that experiments on their own can provide a gold standard for causal discovery. They suggest that causal discovery, whether experimental or observational, depends crucially on the assumptions one makes about the true causal model. As the earlier examples show, assumptions interact with each other and with the available experiments to yield insights about the underlying causal structure. Different sets of assumptions and different sets of experiments result in different degrees of insight and underdetermination, but there is no clear hierarchy either within the set of possible assumptions, or between experiments and assumptions about the model space or parameterization.

3. Interventionism. On the interventionist account of causation, "X is a direct cause of Y with respect to some variable set V if and only if there is a possible intervention on X that will change Y (or the probability distribution of Y) when all other variables in V besides X and Y are held fixed at some value by interventions" (Woodward 2003). The intuition is easy enough: in figure 2, x is a direct cause of z because x and z are dependent in the double-intervention experiment intervening on x and y simultaneously.

According to this definition of a direct cause it is true by definition that N experiments each intervening on $N - 1$ variables are sufficient to identify the causal structure among a set of N variables even when causal sufficiency does not hold. (Above I only discussed necessary conditions.) If each of the N experiments leaves out a different variable from its intervention set, then each experiment can be used to determine the presence of the direct effects from the $N - 1$ intervened variables to the one nonintervened one. Together the experiments determine the entire causal structure.

An interventionist should therefore have no problem with the results discussed so far, since the cases of experimental underdetermination that I have considered were all restricted to experiments intervening on at most $N - 2$ variables. The causal structures could always be distinguished by an experiment intervening on all but one variable.

But there are unusual cases. In appendix C, I provide another parameterization (PM3) for the first causal structure in figure 2 (the one with the extra $x \rightarrow z$ edge). The example and its implications are discussed more thoroughly than can be done here in Eberhardt (2013). PM3 is very similar to PM1 and PM2. In fact, for a passive observation and a single intervention on x, y, or z they all imply the exact same distributions. However, PM3 is also indistinguishable from PM2 for a double-intervention experiment on x and y (and similarly, of course, for all other double-intervention experiments). That is, PM3 and PM2 differ in their causal structure with regard to

https://doi.org/10.1086/673865 Published online by Cambridge University Press

the $x \rightarrow z$ edge but are experimentally indistinguishable for all possible experiments on the observed variables.

In what sense, then, is the direct arrow from $x \rightarrow z$ in PM3 justified? After all, in a double-intervention experiment on x and y, x will appear independent of z. Given Woodward's definition of a direct cause, x is not a direct cause relative to the set of observed variables $\{x, y, z\}$. However, if one included u and v as well, x would become a direct cause of z, since x changes the probability distribution of z in an experiment that changes x and holds y, u, and v fixed.

So, the interventionist can avoid the apparent contradiction. The definition of a direct cause is protected from the implications of PM3 since it is relativized to the set of variables under consideration. But one may find a certain level of discomfort that this interventionist definition permits the possibility that a variable (x here)

(i) is not a direct cause relative to $V = \{x, y, z\}$,
(ii) is not even an indirect cause when y is subject to intervention and $V = \{x, y, z\}$,
(iii) but is a direct cause relative to $V^* = \{x, y, z, u, v\}$.

Unlike PM1, PM3 violates the assumption of faithfulness in the double-intervention distribution when x and y are manipulated simultaneously: in PM3 x is independent of z despite being (directly) causally connected.

Violations of faithfulness have been recognized to cause problems for the interventionist account (Strevens 2008). In particular, when there are two causal pathways between a variable p and a variable q that cancel each other out exactly, then an intervention on p will leave p and q independent despite the (double) causal connection. But this case here is different: in the double-intervention distribution intervening on x and y that is crucial to determining whether x is a direct cause of z, there is only one pathway between x and z. Thus, we are faced here with a violation of faithfulness that does not follow the well-understood case of canceling pathways. But like those cases, it shows that the interventionist account of causation either misses certain causal relations or implicitly depends on additional assumptions about the underlying causal model. The interventionist need not assume faithfulness. As indicated earlier the assumption of linearity guarantees identifiability using only single-intervention experiments even if we do not assume faithfulness. In other words, a linear parameterization of structure 1 cannot be made indistinguishable from a linear parameterization of structure 2.

Part of the appeal of the interventionist account is its sensitivity to the set of variables under consideration when defining causal relations. This helped enormously to disentangle direct from total and contributing causes. Ex-

https://doi.org/10.1086/673865 Published online by Cambridge University Press

amples like PM3 suggest that the relativity may be too general for definitional purposes unless one makes additional assumptions: I may measure one set of variables in an experiment and say there is no causal connection between two variables. You may measure a strict superset of my variables and intervene on a strict superset of my intervened variables and come to the conclusion that the same pair of variables stand in a direct causal relation. Moreover, the claim would hold when all the interventions were successfully surgical, that is, breaking causal connections.

The other part of the interventionist appeal was the apparent independence of the interventionist account from substantive assumptions such as faithfulness that have received little sympathy despite their wide application. This article suggests that you cannot have both.

Appendix A

Theorem: Assuming only causal Markov, faithfulness, and acyclicity, n experiments are in the worst case necessary to discover the causal structure among n variables.

Proof: Suppose that every pair of variables in V is subject to confounding. Consequently, independence tests conditional on any nonintervened variable will always return a dependence, since they open causal connections via the unmeasured variables.

Without loss of generality we can assume that the following about the causal hierarchy over the variables is known:

$$(x_1, x_2) > x_3 > \ldots > x_n.$$

In words: the causal order between x_1 and x_2 is unknown, but they are both higher in the order than any other variable. To satisfy the order, there must (at least) be a path

$$x_3 \rightarrow x_4 \rightarrow \ldots \rightarrow x_{n-1} \rightarrow x_n.$$

Let an experiment $E = (J, U)$ be defined as a partition of the variables in V into a set J and $U = V \backslash J$, where the variables in J are subject to a surgical intervention simultaneously and independently, and the variables in U are not.

Now note the following: the only experiment that establishes whether $x_2 \rightarrow x_1$ are experiments with x_2 in J_1 and x_1 not in J_1. That is, x_2 is subject to an intervention (with possibly other variables) and x_1 is not. Select any one such experiment and call it $E_1 = (J_1, U_1)$.

https://doi.org/10.1086/673865 Published online by Cambridge University Press

Suppose that experiment E_1 showed that x_2 and x_1 were independent, such that the ordering between x_1 and x_2 remains underdetermined.

The only experiment that establishes whether $x_1 \rightarrow x_2$ are experiments E_2 with x_1 in J_2 and x_2 not in J_2.

Experiments E_1 and E_2 resolve the order between x_1 and x_2, suppose without loss of generality that it is $x_1 \rightarrow x_2$. In the worst case this required two experiments.

Now for the remainder: the only experiment that establishes whether $x_1 \rightarrow x_3$ are experiments E_3 with x_1 and x_2 in J_3 and x_3 not in J_3. Note that none of the previous experiments could have been an E_3.

The only experiment that establishes whether $x_1 \rightarrow x_4$ are experiments E_4 with x_1, x_2, x_3 in J_4 and x_4 not in J_4. None of the previous experiments could have been an E_4.

. . .

The only experiment that establishes whether $x_1 \rightarrow x_n$ is an experiment E_n with $x_1, \ldots, x_{n-1}$ in J_n and x_n not in J_n. None of the previous experiments could have been an E_n. It follows that n experiments are in the worst case necessary to discover the causal structure. QED

The above proof shows that in the worst case a sequence of n experiments is necessary that have intervention sets that intervene on at least $n - i$ variables simultaneously for each integer i in $1 < i < n$.

Appendix B

Parameterization PM1 for structure 1 in figure 2 (all variables are binary):

$$p(u = 1) = .5$$

$$p(v = 1) = .5$$

$$p(x = 1|u = 1) = .8$$
$$p(x = 1|u = 0) = .2$$

$$p(y = 1|v = 1, x = 1) = .8$$
$$p(y = 1|v = 1, x = 0) = .8$$
$$p(y = 1|v = 0, x = 1) = .8$$
$$p(y = 1|v = 0, x = 0) = .2$$

$$p(z = 1|u = 1, v = 1, x = 1, y = 1) = .8$$
$$p(z = 1|u = 1, v = 1, x = 1, y = 0) = .8$$
$$p(z = 1|u = 1, v = 1, x = 0, y = 1) = .84$$
$$p(z = 1|u = 1, v = 1, x = 0, y = 0) = .8$$
$$p(z = 1|u = 1, v = 0, x = 1, y = 1) = .8$$
$$p(z = 1|u = 1, v = 0, x = 1, y = 0) = .8$$
$$p(z = 1|u = 1, v = 0, x = 0, y = 1) = .64$$
$$p(z = 1|u = 1, v = 0, x = 0, y = 0) = .8$$
$$p(z = 1|u = 0, v = 1, x = 1, y = 1) = .8$$
$$p(z = 1|u = 0, v = 1, x = 1, y = 0) = .8$$
$$p(z = 1|u = 0, v = 1, x = 0, y = 1) = .79$$
$$p(z = 1|u = 0, v = 1, x = 0, y = 0) = .8$$
$$p(z = 1|u = 0, v = 0, x = 1, y = 1) = .8$$
$$p(z = 1|u = 0, v = 0, x = 1, y = 0) = .2$$
$$p(z = 1|u = 0, v = 0, x = 0, y = 1) = .84$$
$$p(z = 1|u = 0, v = 0, x = 0, y = 0) = .2$$

https://doi.org/10.1086/673865 Published online by Cambridge University Press

Parameterization PM2 for structure 2 in figure 2:

$$
\begin{aligned}
&p(u = 1) = .5 \\
&p(v = 1) = .5 \\
&\\
&p(x = 1|u = 1) = .8 \\
&p(x = 1|u = 0) = .2 \\
&\\
&p(y = 1|v = 1,\ x = 1) = .8 \\
&p(y = 1|v = 1,\ x = 0) = .8 \\
&p(y = 1|v = 0,\ x = 1) = .8 \\
&p(y = 1|v = 0,\ x = 0) = .2
\end{aligned}
\qquad
\begin{aligned}
&p(z = 1|u = 1,\ v = 1,\ y = 1) = .8 \\
&p(z = 1|u = 1,\ v = 1,\ y = 0) = .8 \\
&p(z = 1|u = 1,\ v = 0,\ y = 1) = .8 \\
&p(z = 1|u = 1,\ v = 0,\ y = 0) = .8 \\
&p(z = 1|u = 0,\ v = 1,\ y = 1) = .8 \\
&p(z = 1|u = 0,\ v = 1,\ y = 0) = .8 \\
&p(z = 1|u = 0,\ v = 0,\ y = 1) = .8 \\
&p(z = 1|u = 0,\ v = 0,\ y = 0) = .2
\end{aligned}
$$

Passive observational distribution:

$$\text{PM1: } P(X, Y, Z) = \sum_{uv} P(U)P(V)P(X|U)P(Y|V, X)P(Z|U, V, X, Y)$$

$$\text{PM2: } P(X, Y, Z) = \sum_{uv} P(U)P(V)P(X|U)P(Y|V, X)P(Z|U, V, Y)$$

Experimental distribution when x is subject to an intervention (I write $P(A|B||B)$ to mean the conditional probability of A given B in an experiment where B has been subject to a surgical intervention):

$$\text{PM1: } P(Y, Z|X||X) = \sum_{uv} P(U)P(V)P(Y|V, X)P(Z|U, V, X, Y)$$

$$\text{PM2: } P(Y, Z|X||X) = \sum_{uv} P(U)P(V)P(Y|V, X)P(Z|U, V, Y)$$

Experimental distribution when y is subject to an intervention:

$$\text{PM1: } P(X, Z|Y||Y) = \sum_{uv} P(U)P(V)P(X|U)P(Z|U, V, X, Y)$$

$$\text{PM2: } P(X, Z|Y||Y) = \sum_{uv} P(U)P(V)P(X|U)P(Z|U, V, Y)$$

Experimental distribution when z is subject to an intervention:

$$\text{PM1: } P(X, Y|Z||Z) = \sum_{uv} P(U)P(V)P(X|U)P(Y|V, X)$$

$$\text{PM2: } P(X, Y|Z||Z) = \sum_{uv} P(U)P(V)P(X|U)P(Y|V, X)$$

By substituting the terms of PM1 and PM2 in the above equations it can be verified that PM1 and PM2 have identical passive observational and single-intervention distributions but that they differ for the following double-intervention distribution on x and y.

https://doi.org/10.1086/673865 Published online by Cambridge University Press

Experimental distribution when x and y are subject to an intervention:

$$\text{PM1: } P(Z|X,Y||X,Y) = \sum_{uv} P(U)P(V)P(Z|U,V,X,Y)$$

$$\text{PM2: } P(Z|X,Y||X,Y) = \sum_{uv} P(U)P(V)P(Z|U,V,Y)$$

PM1 and PM2 (unsurprisingly) have identical distributions for the other two double-intervention distributions, since the $x \rightarrow z$ edge is broken and the remaining parameters are identical in the parameterizations:

Experimental distribution when x and z are subject to an intervention:

$$\text{PM1: } P(Y|X,Z||X,Z) = \sum_{v} P(V)P(Y|V,X)$$

$$\text{PM2: } P(Y|X,Z||X,Z) = \sum_{v} P(V)P(Y|V,X)$$

Experimental distribution when y and z are subject to an intervention:

$$\text{PM1: } P(X|Y,Z||Y,Z) = \sum_{u} P(U)P(X|U)$$

$$\text{PM2: } P(X|Y,Z||Y,Z) = \sum_{u} P(U)P(X|U)$$

Appendix C

Parameterization PM3 for structure 1 in figure 2:

$p(u = 1) = .5$

$p(v = 1) = .5$

$p(x = 1|u = 1) = .8$
$p(x = 1|u = 0) = .2$

$p(y = 1|v = 1,\ x = 1) = .8$
$p(y = 1|v = 1,\ x = 0) = .8$
$p(y = 1|v = 0,\ x = 1) = .8$
$p(y = 1|v = 0,\ x = 0) = .2$

$\boldsymbol{p(z = 1|u = 1, v = 1, x = 1, y = 1) = .825}$
$p(z = 1|u = 1,\ v = 1,\ x = 1,\ y = 0) = .8$
$\boldsymbol{p(z = 1|u = 1, v = 1, x = 0, y = 1) = .8}$
$p(z = 1|u = 1,\ v = 1,\ x = 0,\ y = 0) = .8$
$p(z = 1|u = 1,\ v = 0,\ x = 1,\ y = 1) = .775$
$p(z = 1|u = 1,\ v = 0,\ x = 1,\ y = 0) = .8$
$p(z = 1|u = 1,\ v = 0,\ x = 0,\ y = 1) = .8$
$p(z = 1|u = 1,\ v = 0,\ x = 0,\ y = 0) = .8$
$p(z = 1|u = 0,\ v = 1,\ x = 1,\ y = 1) = .7$
$p(z = 1|u = 0,\ v = 1,\ x = 1,\ y = 0) = .8$
$p(z = 1|u = 0,\ v = 1,\ x = 0,\ y = 1) = .8$
$p(z = 1|u = 0,\ v = 1,\ x = 0,\ y = 0) = .8$
$p(z = 1|u = 0,\ v = 0,\ x = 1,\ y = 1) = .9$
$p(z = 1|u = 0,\ v = 0,\ x = 1,\ y = 0) = .2$
$p(z = 1|u = 0,\ v = 0,\ x = 0,\ y = 1) = .8$
$p(z = 1|u = 0,\ v = 0,\ x = 0,\ y = 0) = .2$

Substituting the parameters of PM3 in the equations for the passive observational or any experimental distributions of PM1 in appendix 2, it can be

https://doi.org/10.1086/673865 Published online by Cambridge University Press

verified that PM2 and PM3 are experimentally indistinguishable for all possible experiments on $\{x, y, z\}$. Nevertheless, it should be evident that in an experiment intervening on x, y, u, and v, the difference between the bold font parameters will indicate that x is a direct cause of z.

REFERENCES

Eberhardt, Frederick. 2013. "Direct Causes." Unpublished manuscript, PhilSci Archive. http://philsci-archive.pitt.edu/9502/.

Eberhardt, Frederick, Clark Glymour, and Richard Scheines. 2005. "On the Number of Experiments Sufficient and in the Worst Case Necessary to Identify All Causal Relations among *n* Variables." In *Proceedings of the 21st Conference on Uncertainty and Artificial Intelligence*, ed. Fahiem Bacchus and Tommi Jaakkola, 178–84. Arlington, VA: AUAI.

Eberhardt, Frederick, Patrik O. Hoyer, and Richard Scheines. 2010. "Combining Experiments to Discover Linear Cyclic Models with Latent Variables." AISTATS 2010 conference proceedings. *Journal of Machine Learning Research: Workshop and Conference Proceedings*. Vol. 9. http://machinelearning.wustl.edu/mlpapers/papers/AISTATS2010_EberhardtHS10.

Fisher, Ronald A. 1935. *The Design of Experiments*. New York: Hafner.

Geiger, Dan, Thomas Verma, and Judea Pearl. 1990. "Identifying Independence in Bayesian Networks." *Networks* 20:507–34.

Hyttinen, Antti, Frederick Eberhardt, and Patrik O. Hoyer. 2011. "Noisy-or Models with Latent Confounding." In *Proceedings of the 27th Conference on Uncertainty and Artificial Intelligence*, ed. Fabio G. Cozman and Avi Pfeffer. Corvallis, OR: AUAI.

Pearl, Judea. 2000. *Causality.* Oxford: Oxford University Press.

Shimizu, Shohei, Patrik O. Hoyer, Aapo Hyvarinen, and Antti J. Kerminen. 2006. "A Linear Non-Gaussian Acyclic Model for Causal Discovery." *Journal of Machine Learning Research* 7:2003–30.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction and Search*. Cambridge, MA: MIT Press.

Strevens, Michael. 2008. "Comments on Woodward, *Making Things Happen*." *Philosophy and Phenomenological Research* 77:171–92.

Woodward, James. 2003. *Making Things Happen*. Oxford: Oxford University Press.

https://doi.org/10.1086/673865 Published online by Cambridge University Press