Probabilistic Causation and the Markov Condition

J. Dmitri Gallow†

Abstract

On one understanding of probabilistic causation, the probability attaches to the effect—c probabilistically caused e iff c deterministically caused the probability of e. On the other understanding, the probability attaches to the causal relation—c probabilistically caused e iff c caused e, though there was some probability that it wouldn’t. Hausman & Woodward (1999, 2004) appeal to the former conception of probabilistic causation to show that the Causal Markov Condition will be satisfied when causation is probabilistic. I demonstrate that, pace Hausman & Woodward, this conception of probabilistic causation is incapable of establishing the Causal Markov Condition except by fiat. I then demonstrate that, on the latter conception of probabilistic causation, there are violations of the Markov Condition; however, there is a general locality constraint which is sufficient to restore the Markov Condition.

1 Introduction

More and more, applied work in fields like Economics, Sociology, and Epidemiology is utilizing the methods of causal modeling and causal inference developed and championed by Spirtes et al. (2000) and Pearl (2000). These methods rely crucially on an assumption known as the Causal Markov Condition (cmc). The cmc claims that, from the world’s causal structure, we can infer something about the objective probability function, viz., that it satisfies the Markov Condition with respect to that causal structure. Roughly, the Markov Condition is the condition that the value of a variable

Draft of September 22, 2014; please cite and circulate widely
✉: jdmitrig@nyu.edu
† I am indebted to David Baker, Gordon Belot, Clark Glymour, Daniel Hausman, Alistair Isaac, James Joyce, Jonathan Shaheen, Daniel J. Singer, Naftali Weinberger, and the philosophy of science reading group at the University of Wisconsin–Madison for comments on earlier versions of this paper
is independent of the values of any variables it doesn’t causally influence, once we know the values of the variables that immediately causally influence it.

Most everyone seems to agree that, when causation is deterministic, the CMC is correct. Cartwright (1993, 2002, 2006) claims that, when causation is probabilistic, the CMC will fail. If we think that most of the causal relations studied in fields like Economics, Sociology, and Epidemiology are probabilistic causal relations, then this would give us serious reason to worry about the use of Spirtes et al.’s and Pearl’s techniques of causal inference in those fields.

Hausman & Woodward (1999) offer a defense of the probabilistic CMC. This defense presupposes a certain conception of probabilistic causation. On this conception, the probability attaches to the effect. What it is for \( c \) to probabilistically cause \( e \) is for \( c \) to cause the probability of \( e \). Probabilistic causation is just causation of probabilities.

\[ c \text{ probabilistically caused } e \iff c \text{ caused } \Pr(e) \land \Pr(e) < 1 \]

Call this the thesis of probabilistic effects. Hausman & Woodward claim that if we incorporate the thesis of probabilistic effects into the causal modeling framework, then it follows from the thesis that deterministic causal structures satisfy the Markov Condition—the deterministic CMC—that probabilistic causal structures will satisfy the Markov Condition as well—the probabilistic CMC.

Hausman & Woodward are incorrect. The thesis of probabilistic effects, incorporated straightforwardly into the causal modeling framework, entails the negation of the probabilistic Markov Condition. The only way to guarantee the probabilistic Markov Condition in models built around the thesis of probabilistic effects is to write it in by hand. The thesis of probabilistic effects, therefore, cannot offer a vindication of the probabilistic CMC. There is, however, another way of thinking about probabilistic causation. On this view, the probability attaches to the causal relation. What it is for \( c \) to probabilistically cause \( e \) is for \( c \) to cause \( e \), even though there was some probability that it wouldn’t.

\[ c \text{ probabilistically caused } e \iff c \text{ caused } e \land \Pr(c \text{ cause } e) < 1 \]

Call this the thesis of probabilistic causation. The thesis of probabilistic causation succeeds where the thesis of probabilistic effects fails. By incorporating the thesis of probabilistic causation into the causal modeling framework, we can show that—though there are some counterexamples to the Markov
Condition—when a general locality constraint I dub *structural independence* is satisfied, the Markov Condition follows, even when causation is probabilistic.

In §2 below, I’ll introduce the causal modeling techniques of Pearl and Spirtes et al. In §3, I’ll introduce Cartwright’s arguments against the probabilistic cmc, as well as Hausman & Woodward’s response to these arguments. In §4, I’ll explain how Hausman & Woodward use the thesis of probabilistic effects to argue for the probabilistic cmc, and why this strategy fails. I’ll then show, in §5, how to construct models of probabilistic causal structure based upon the thesis of probabilistic causation. I’ll show that, though these models fail to validate the Markov Condition in general, if a model satisfies the principle I’ll call *structural independence*, then it will also satisfy the Markov Condition. I’ll conclude by drawing some lessons about the debate over the probabilistic cmc.

## 2 Causal Models

A causal model \( \mathcal{M} \) is a 4-tuple \( \langle \mathcal{U}, \mathcal{V}, \mathcal{E}, \mathcal{P} \rangle \) of a set \( \mathcal{U} = \{U_1, U_2, ..., U_N\} \) of *exogenous* variables, a set \( \mathcal{V} = \{V_1, V_2, ..., V_M\} \) of *endogenous* variables, a set \( \mathcal{E} = \{\phi_{V_1}, \phi_{V_2}, ..., \phi_{V_M}\} \) of structural equations (one for each endogenous variable), and a probability function \( \mathcal{P} \), defined over the values of the variables in \( \mathcal{U} \).

The structural equations in \( \mathcal{E} \) establish functional determination relations between the variables in \( \mathcal{U} \cup \mathcal{V} \). For instance, consider the causal model \( \mathcal{M}_1 = \langle \mathcal{U}_1, \mathcal{V}_1, \mathcal{E}_1, \mathcal{P}_1 \rangle \), where \( \mathcal{U}_1 = \{U_X, U_Y\} \), \( \mathcal{V}_1 = \{X, Y, Z\} \), and

\[
\mathcal{E}_1 = \left\{ \begin{array}{l}
X := \phi_X(U_X) \\
Y := \phi_Y(X, U_Y) \\
Z := \phi_Z(X, Y)
\end{array} \right\}.
\]

(Don’t worry about the probability function for now.) In this model, \( \mathcal{E}_1 \) tells us that the value of \( X \) is determined by the value of \( U_X \), the value of \( Y \) is determined by the values of \( X \) and \( U_Y \), and the value of \( Z \) is determined by the values of \( X \) and \( Y \). What makes these equations *structural* is that they cannot be re-arranged as an ordinary equation could. Ordinarily, so long as \( \phi_X \) is bijective, we would be able to rewrite \( X = \phi_X(U_X) \) as \( U_X = \phi_X^{-1}(X) \).

---

1 A word on notation: throughout, I will use uppercase Latin letters, usually from the end of the alphabet (...X, Y, Z), to represent variables, and the corresponding lowercase letters (...x, y, z) to stand for the values of those variables. Boldface Latin letters (...X, Y, Z) will be use to refer to sets of variables. Functions will be denoted using \( \phi \), subscripted with variables. At times, I will use the function name alone to refer to the entire structural equation.
We might then be tempted to conclude that the value of $U_X$ is determined by the value of $X$ just as much as the value of $X$ is determined by the value of $U_X$. Bijective functional relationships are symmetrical. Structural equations, on the other hand, are asymmetrical. It matters which variable is to the left of the equals sign. $'X := \phi_X(U_X)'$ says that $X$ is determined by $U_X$ in a way that $U_X$ is not determined by $X$. (To emphasize this asymmetry, I am using ‘$\equiv$’ to distinguish that asymmetrical relation from the symmetrical $=\)."

Given a causal model $(U, V, E, P)$, we can generate a causal graph as follows. For any variables $X, Y \in U \cup V$, if $Y$ appears on the right-hand-side of $X$’s structural equation $\phi_X \in E$, then place an arrow between $X$ and $Y$, with its tail at $Y$ and its head at $X$. For instance, the model $M_1$ generates the graph shown in figure 1(a). That graph tells us that $X$ is a function of $U_X$ alone, that $Y$ is a function of $X$ and $U_Y$ alone, and that $Z$ is a function of $X$ and $Y$ alone. Throughout, I’ll denote the exogenous variable associated with the endogenous variable $V$ with ‘$U_V’”, and dash the arrows emanating from exogenous variables in the graph. This latter convention will allow me to leave the exogenous variables out of the graph entirely when they are not important for my purposes. Thus, both the graph in figure 1(a) and the graph in figure 1(b) could be used to represent the model $M_1$.

Call any sequence of arrows along the graph (e.g., $Z \leftarrow X \rightarrow Y$) a path. If all the arrows in a path are oriented tail-to-tip (as in $X \rightarrow Y \rightarrow Z$), call it a directed path. We can use the language of genealogy to talk about the structural relationships amongst these variables. If a variable lies to the right of $X$ along some directed path, call it a descendant of $X$. Also, let’s stipulate that every variable is own of its own descendants. If there’s an arrow with its tail at $Y$ and its head at $X$, call $Y$ a parent of $X$. Denote the set of a variable $X$’s endogenous parents with ‘$\text{PA}(X)$’, and denote the set of $X$’s descendants with ‘$\text{DE}(X)$’. In the causal graph in figure 1(a), for instance, $\text{PA}(Y) = \{X\}$ and $\text{DE}(Y) = \{Y, Z\}$.

It is standard to restrict attention to those causal models which are acyclic.

**Acyclicity.** For every $V \in V$, $\text{PA}(V) \cap \text{DE}(V) = \emptyset$. 

---

**Figure 1**

\[ U_X \rightarrow \rightarrow X \rightarrow Y \leftarrow \rightarrow U_Y \]

(a)

\[ X \rightarrow \rightarrow Y \leftarrow \rightarrow U_Y \]

(b)
A causal model which is acyclic is just one whose associated causal graph does not have any loops, or directed paths which begin and end with the same variable.

Though these models include a probability function, $\mathcal{P}$, this does not mean that they model probabilistic causation. On the contrary, these causal models presuppose causal determinism—they assume that the value of every endogenous variable is entirely determined by the values of its parent variables in the model. The structural equations in $\mathcal{E}$ tell us precisely how they are determined by those variables. The exogenous variables, however, are not fixed by the model. Rather, we just have a probability distribution, $\mathcal{P}$, over the possible values of these variables. Once the values of the exogenous variables become a matter of probability, so too do those of all the endogenous variables, since these values are determined by those of the exogenous variables, via the deterministic functional relationships given by the structural equations in $\mathcal{E}$. So a joint probability distribution over the exogenous variables will induce a joint probability distribution over the endogenous variables as well, if we assume that the probability that a variable $V$ takes on the value $v$ is equal to the probability that $V$'s parents take on the values $\phi^{-1}_V(V = v)$,

$$\mathcal{P}(V = v) = \mathcal{P}(\phi^{-1}_V(V = v))$$

(1)

where '$\phi^{-1}_V(V = v)$' is the pre-image of $V = v$, under the function $\phi_V$.

If $X$ and $Y$ are independent according to the probability function $\mathcal{P}$, then I will write ‘$X \perp \mathcal{P} Y$’.

$$X \perp \mathcal{P} Y \equiv \mathcal{P}(X, Y) = \mathcal{P}(X) \cdot \mathcal{P}(Y)$$

Similarly, if $X$ and $Y$ are independent conditional on $W$, then I will write ‘$X \perp \mathcal{P} Y \mid W$’.

$$X \perp \mathcal{P} Y \mid W \equiv \mathcal{P}(X, Y \mid W) = \mathcal{P}(X \mid W) \cdot \mathcal{P}(Y \mid W)$$

When the probability function in question is clear from context, I will suppress the subscripted ‘$\mathcal{P}$’.

It is standard to restrict attention to those causal models in which no two endogenous variables have the same exogenous variable as a parent, and whose associated probability functions deem all (disjoint sets of) exogenous

---

2 Throughout, expressions like ‘$\mathcal{P}(X, Y) = \mathcal{P}(X) \cdot \mathcal{P}(Y)$’, mean that $\forall x \forall y \mathcal{P}(X = x, Y = y) = \mathcal{P}(X = x) \cdot \mathcal{P}(Y = y)$. Also, I'll write ‘$X = x, Y = y$’, or ‘$X, Y$’, for the conjunction of $X = x$ and $Y = y$. 

---
variables to be independent of one another.

**Exogenous Independence.** Every exogenous variable \( U \in \mathcal{U} \) has at most one child, and for any two disjoint sets of exogenous variables \( U_1 \) and \( U_2 \),

\[
U_1 \perp U_2
\]

Call any model which satisfies **Acyclicity** and **Exogenous Independence** a *Markovian* model.

### 2.1 The Causal Markov Condition

The Markov Condition is a condition on the probability function defined over the variables appearing in a causal model. It says that, once we know the values of the parents of all of the variables in \( X \), knowledge of the values of any other variables in the model, other than the descendants of the variables in \( X \), will not give us any information about the values of the variables in \( X \). That is, \( X \)'s parents screen it off from its non-descendants.\(^3\)

**Markov Condition.** For any \( X, Y \subseteq \mathcal{V} \) such that \( Y \cap \text{DE}(X) = \emptyset \),

\[
X \perp Y \mid PA(X)
\]

For instance, if the model shown in figure 2 satisfies the Markov Condition, then \( Y \) is independent of \( W \) conditional on \( X \) (\( Y \perp W \mid X \)), and \( Z \) is independent of \( X \) conditional on \( W \) and \( Y \) (\( Z \perp X \mid W, Y \)).

It is a theorem that every Markovian causal model satisfies the Markov Condition.

**Theorem 1.** If a model \( \mathcal{M} = \langle \mathcal{U}, \mathcal{V}, \mathcal{E}, \mathcal{P} \rangle \) is Markovian, then for any \( X, Y \subseteq \mathcal{V} \) such that \( Y \cap \text{DE}(X) = \emptyset \),

\[
X \perp Y \mid PA(X)
\]

\(^3\) Above, I only defined \( PA(\cdot) \) and \( \text{DE}(\cdot) \) for individual variables. Let \( PA(X) \) be \( \bigcup_i PA(X_i) \), for every \( X_i \in X \). Likewise for \( \text{DE}(X) \).
This theorem tells us one of the properties of a certain family of mathematical objects, the Markovian causal models. Following Hitchcock (2010), we should distinguish Theorem 1 from what has been called the Causal Markov Condition. The former is just a theorem about a family of mathematical models; the latter is a claim about the objective probability function and the world’s actual causal structure. The Causal Markov Condition says that Markovian causal models serve as a good characterization of the relationship between causal structure and objective probability, that the actual world’s causal structure obeys the Markov Condition.

**Causal Markov Condition** (cmc). For any causal graph relating the variables in an appropriate variable set \( V \), if the causal graph correctly describes the causal structure of the variables in \( V \) in the actual world, then, for every \( X, Y \subseteq V \) such that \( Y \cap \text{DE}(X) = \emptyset \),

\[
X \perp\!
\!
\perp_{\text{Pr}} Y \mid \text{PA}(X)
\]

where \( \text{Pr} \) is the objective probability function.

In the next section, I’ll introduce a purported counterexample to the Causal Markov Condition in cases involving probabilistic causation due to Cartwright, as well as a response to the counterexample offered by Hausman & Woodward. In §4, I’ll consider an argument from Hausman & Woodward that the Markov Condition will be satisfied in cases involving probabilistic causation which relies on the assumption that probabilistic causation is deterministic causation of probabilities—what I earlier called the thesis of probabilistic effects. I’ll show that this assumption, imported naturally into the structural equations framework, entails that the Markov Condition is not, in general, satisfied. The only way to guarantee the Markov Condition, in this framework, is to write it in by hand. This will motivate the models I will introduce in §5.1, which are built around the thesis of probabilistic causation—the thesis that, if \( c \) caused \( e \) probabilistically, then \( c \) caused \( e \), though there was some probability it wouldn’t.

### 3 Against the Probabilistic Causal Markov Condition

In her 1993, Nancy Cartwright tells us the following story: There is a certain chemical factory, called ‘Cheap-but-Dirty’, which produces a water-treatment chemical for the city’s water supply. Cheap-but-Dirty is, however, unscrupulous in its methods. Rather than just treating the water every day, the factory activates a purely chance process which has only an 80% chance of resulting in the water-treatment chemical being produced that day. Thus,
20% of the time, the water simply goes untreated. Given their low price, the city is willing to tolerate its water being dirty 20% of the time. However, they draw the line at the factory’s effluvia, which is polluting the local rivers. This effluvia is produced whenever the water treatment chemical is.

The city is about to pull the plug on Cheap-but-Dirty when its owners—well-trained in contemporary causal modeling techniques—object that it’s not their factory which is to blame, but rather the water-treatment chemical itself which causes the effluvia. Their argument goes as follows: the city alleges that the factory is causing both the chemical and the effluvia. However, if that were so, then the correct causal graph would be the one shown in figure 3. There, $F$, $C$, and $E$ are binary variables representing the state of the factory, the production of the chemical, and the presence of the effluvia, respectively. ($F = 1$ if the factory is running and 0 otherwise; $C = 1$ if the chemical is produced and 0 otherwise; $E = 1$ if the rivers are polluted with effluvia and 0 otherwise.) But if this is the correct causal structure of $F$, $C$, and $E$, then $\text{cmc}$ tells us that $C$ should be independent of $E$, conditional on $F$. Cheap-but-Dirty’s owners then produce statistics demonstrating that\footnote{Throughout, when I am dealing with a binary variable, $V$, I will use ‘$V$’ to denote $V = 1$ and ‘$\overline{V}$’ to denote $V = 0$.}

$$\Pr(E \mid F) = 0.8$$

while

$$\Pr(E \mid F, C) = 1$$

Therefore, $C \not\perp_{\Pr} E \mid F$. So figure 3 cannot represent the true causal structure of $F$, $C$, and $E$. Cheap-but-Dirty’s owners then suggest the actual causal structure between these variables is the one shown in figure 4. For the only non-trivial independence relation the $\text{cmc}$ entails if this is the correct causal structure is $F \perp_{\Pr} E \mid C$, and $F$ and $E$ are independent once we condition
on C:

\[ \Pr(E \mid C, F) = \Pr(E \mid C) = 1 \]

So, they suggest, we shouldn’t conclude that it’s the substandard factory which is causing all the harmful effluvia; rather, we should conclude that the water-treatment chemical has the awful side effect of polluting the rivers.

Clearly, Cheap-but-Dirty’s owners have just pulled a cheap and dirty trick. The reason that \( \Pr(E \mid F) = 0.8 \) while \( \Pr(E \mid F, C) = 1 \) is just that the factory produces the chemical when and only when it produces the effluvia, and both of these occur about 80% of the days that the factory is running. So, even if we know that the factory is running, learning whether the water treatment chemical is in the water supply tells us whether the factory’s purely probabilistic process has actually resulted in the production of the chemical; and, given that the factory’s probabilistic process produces the chemical when and only when it produces the effluvia, learning whether the chemical has been produced tells us whether there is effluvia in the river.

But none of this means that the chemical is causing the effluvia. The factory is causing the effluvia. Figure 3 is the correct causal diagram. What has gone wrong, Cartwright maintains, is that, when a cause has joint probabilistic effects, and one effect is caused whenever the other is, there will be failures of the cmc. Cartwright claims that systems with this causal structure are rife in the sorts of macroscopic probabilistic systems we regularly encounter. If Daniel has a cold, and I accidentally drink from his cup, this may or may not cause me to get a sore throat, and it may or may not cause me to get a runny nose. However, if it does cause a sore throat, then it is significantly more likely to cause a runny nose. So then, even though my runny nose and my sore throat are both probabilistic effects of drinking from Daniel’s cup, they are not independent, conditional on their common cause.

Hausman & Woodward (1999) provide two responses to Cartwright’s counterexample. The first is to claim that there is something illicit about the variables being used to model the case. The second is to claim that there is something illicit about the variable set. I’ll discuss the first of these responses here, putting off my discussion of the second response to the conclusion.

3.1 Inappropriate Variables

There are a few restrictions which must be placed on the variables in \( V \) in order for the cmc to not fall to trivial counterexamples. For instance, the variables in \( V \) must be distinct, in the sense of Lewis (1986)—there must not be any purely logical connections between the values of the variables in \( V \). For instance, if \( X \) is annual rainfall in Atlanta and \( Y \) is annual precipitation
in Atlanta, then $X$ and $Y$ will be correlated even after we condition on all of the causes of $Y$. That’s because the value of rainfall places a lower bound on the value of precipitation and the value of precipitation places an upper bound on the value of rainfall. But this is a logical, and not a causal relationship between the two variables. So variable sets like these have to be ruled out right off the bat.

There is another restriction which, it is sometimes assumed, must be placed on $V$: the variables in $V$ cannot be too coarsely-grained. Consider the following example from Salmon (1984): I hit the cue ball. The cue ball collides with the eight ball, which sinks in the corner pocket. The cue ball then scratches in the adjacent corner pocket. Given the position of the cue and eight ball on the table, the eight ball will sink just in case the cue ball scratches. So, if we look at the variable set $\{H, S_8, S_C\}$—where $H$ is 1 if the cue ball hits the eight ball and 0 otherwise, $S_8$ is 1 if the eight ball sinks in the corner pocket and is 0 otherwise, and $S_C$ is 1 if the cue ball scratches and is 0 otherwise—then, even though the collision caused the eight ball to sink and caused the cue ball to scratch, and the eight ball’s sinking did not cause the cue ball’s scratching, nor did the cue ball’s scratching cause the eight ball’s sinking, $\Pr(S_C \mid H) \neq \Pr(S_C \mid S_8, H)$. Even after we take into account that the cue ball hit the eight ball, learning that the eight ball sunk tells us that the cue ball must have scratched. For, given the placement of the balls on the table, it was guaranteed that the the cue ball would scratch if the eight ball sunk, but it’s not the case that the cue ball was guaranteed to scratch if it hit the eight ball. So $\Pr(S_C \mid H) < 1$, and $\Pr(S_C \mid S_8, H) = 1$, so $S_8 \not\perp \!\!\!\perp \Pr, S_C \mid H$. So this variable set violates the cmc.

Notice, however, that the variable set $\{H^+, S_8, S_C\}$—where $H^+$ is a binary variable which takes the value 1 iff the cue ball hits the eight ball at just the right angle to send the eight and the cue balls off to their respective pockets—will satisfy the cmc. For, once we know that $H^+ = 1$, we know that both $S_8$ and $S_C$ are 1 as well. So $S_8 \perp \!\!\!\perp \Pr, S_C \mid H^+$. So, some have concluded, in order to satisfy the Markov Condition, we must make sure that we are looking at variables which are sufficiently fine-grained.

Similarly, Hausman & Woodward (1999) argue, the variable $F$, which simply tells us whether the factory has been turned on, is not sufficiently fine-grained. The solution is to replace it with a variable, $F^+$, which takes the value 1 if the factory is turned on and its probabilistic process leads it to produce the chemical, and is 0 if the factory isn’t turned on or is turned on and its probabilistic process doesn’t lead it to produce the chemical. Then,

---

5 Cf., e.g., Hausman & Woodward (1999, p. 528–29) and Spirtes et al. (2000, p. 37).
This can’t be quite right. While there’s a sense in which $H^+$ and $F^+$ are more finely grained than $H$ and $F$—they map fewer possibilities to 1 than $H$ and $F$—there’s also a good sense in which they are more coarsely grained than $H$ and $F$—they map more possibilities to 0 than $H$ and $F$. The variables $H^+$ and $F^+$ aren’t strictly more informative than $H$ and $F$—rather, they simply provide a different kind of information, and that kind of information is the kind of information we need in order to restore the Markov Condition. But now all of this can begin to seem a bit delphic. What singles out the information provided by $H^+$ and $F^+$ as being the right kind of information? More pressing: how are practicing researchers to divine which variables are the right ones to be using? Does the cmc amount to nothing more than this—that there are some variables such that the objective probability distribution over those variables will satisfy the Markov Condition?

In fact, I think there is much more to be said for the cmc than this. In my view, the right lesson to draw here isn’t that $H^+$ and $F^+$ restore the Markov Condition because they are more or less fine-grained. $H^+$ and $F^+$ restore the Markov Condition because they provide the kind of information that allows us to write down true structural equations. $S_8 := H^+$ and $S_C := H^+$ are both true structural equations, whereas $S_8 := H$ and $S_C := H$ are not. It’s not true that, if the cue ball hits the eight ball, then the eight ball will sink; nor that, if the cue ball hits the eight ball, that the cue ball will scratch. This, it seems to me, is the difference that makes a difference between $H^+$ and $H$—not that one is more or less fine-grained, or more or less informative, than the other. The difference is that there is a strict structural determination relationship between the values of the variables $S_8$ and $H^+$, while there is no strict structural determination between the values of the variables $S_8$ and $H$. This gives us a principled grounds for rejecting the variable $H$ in favor of the variable $H^+$—one of them is the parent of $S_8$ in a correct causal model and the other is not.

None of this, however, seems to me to provide an adequate response to Cartwright. To say that we can write down a set of true structural equations relating the values of the variables is just to say that we are dealing with a causally deterministic system. But Cartwright readily concedes that the Markov Condition holds when causation is deterministic. What she was attempting to provide was a case in which the structural determination relations were probabilistic. Simply maintaining that, since there are true, deterministic structural equations to be found in Chemical Factory, we must pay attention to those, and not the probabilistic structural equations to which
Cartwright was drawing our attention, is just refusing to play along.

I take it that part of Cartwright’s point is that the variables studied by
the special sciences are much more like $H$ and $F$ than $H^+$ and $F^+$. The kinds
of hypotheses being considered by epidemiologists are hypotheses about
the relationship between variables like smoking and lung cancer—and there need
not be, and probably is not, a deterministic structural relationship between
these variables. There may be some complicated variables describing the pre-
cise amount that a person smokes and the state of the cancerous cells in that
person’s body such that we can write down a true structural equation relat-
ing the values of those variables. But we should not pretend that these are
the kinds of causal hypotheses that practicing researchers are formulating and
testing with the cmc.

4 For the Probabilistic Causal Markov Condition

Hausman & Woodward (1999) provide an independent argument for the
conclusion that the cmc will be satisfied by probabilistic causal systems. This
argument appeals to the thesis of probabilistic effects. They argue that if we
grant this thesis and we think that the cmc holds for deterministic causa-
tion (as Cartwright does), then we should think that the cmc holds for
probabilistic causation as well.

Hausman & Woodward argue that the right way to incorporate the the-
sis of probabilistic effects into the causal modeling framework is to say that,
when $X$ is a probabilistic causal parent of $Y$, $X$ is a deterministic causal par-
ent of $Pr_Y$, where $Pr_Y$ is a variable ranging over probability functions defined
over the values of $Y$. That is, a probabilistic model within which $X \in \text{PA}(Y)$
is equivalent to a deterministic causal model within which $X \in \text{PA}(Pr_Y)$.

Probabilistic Effects: A system in which $X$ is a probabilistic causal
parent of $Y$ is properly modeled by a deterministic causal model
in which $X$ is a causal parent of $Pr_Y$.

What about $Y$? Does $Pr_Y$ causally influence $Y$ in turn? As Cartwright
(2002) points out, if $Pr_Y$ causally influences $Y$, then it must do so probabilis-
tically; and according to Probabilistic Effects, for that to be the case is for $Pr_Y$
to causally influence $Pr_Y$. But if $Pr_Y$ is a parent of $Pr_Y$, what need is there of
$X$? Fortunately, Hausman & Woodward do not suggest that $Pr_Y$ is a causal
parent of $Y$. Though the view is not developed in their article, the curious
reader is directed to chapter 9 of Hausman (1998). There, we are told that,
on this conception of probabilistic causation, $Y$ is simply uncaused. Then,
the proper deterministic causal graph of a probabilistic structural determination relation between \( X \) and \( Y \) is the one shown in figure 5.

With \textit{Probabilistic Effects} in tow, the argument for the probabilistic cmc proceeds as follows. For any \( X, Y \subseteq \mathcal{V} \),

1. If no \( Y \in Y \) is a probabilistic descendant of any \( X \in X \), then for no \( Y \in Y \) is \( \text{Pr}_Y \) a deterministic causal descendant of any \( X \in X \), i.e., \( \text{Pr}_Y \cap \text{DE}(X) = \emptyset \), where \( \text{Pr}_Y \overset{\text{def}}{=} \{ \text{Pr}_Y : Y \in Y \} \)

2. If \( \text{Pr}_Y \cap \text{DE}(X) = \emptyset \), then \( X \perp \text{Pr}_Y | \text{PA}(X) \)

3. If \( X \perp \text{Pr}_Y | \text{PA}(X) \), then \( X \perp Y | \text{PA}(X) \)

4. So, if no \( Y \in Y \) is a probabilistic causal descendant of any \( X \in X \), then \( X \perp Y | \text{PA}(X) \)

1 follows from \textit{Probabilistic Effects}, and 2 is meant to follow from the deterministic cmc.

\textsc{Cartwright} (\textit{2002}) responds to this argument by alleging that 3 can be false in cases of probabilistic causation. She illustrates this by presenting a variant of her Chemical Factory case in which 3 is violated. Whatever we want to say about that case, if we accept \textit{Probabilistic Effects}, then 2 is also false in models containing two successive probabilistic structural determination relations. For suppose that \( X \) is a probabilistic parent of \( W \), and \( W \) is a probabilistic parent of \( Y \). \textit{Probabilistic Effects} tells us that this should be modeled with a deterministic causal model within which \( X \) is a causal parent of \( \text{Pr}_W \), \( W \) is parentless but is a causal parent of \( \text{Pr}_Y \), and \( Y \) is parentless. That is, it tells us that the causal graph shown in figure 6 is correct. In this graph, \( \text{Pr}_Y \) is not one of the deterministic causal descendants of \( X \). However,
it will be false, in general, that $X \perp \! \! \! \perp \Pr_Y \mid \mathbf{PA}(X)$. $Y$ is, after all, one of the probabilistic causal descendants of $X$. In any model like this, $2$ will be false.

The premise is not merely false. It is false in a particularly bad way. It is false in a way that does more than dash Hausman & Woodward’s argument for the probabilistic cmc. It is false in a way that demonstrates that Probabilistic Effects entails, on its own, that the Markov Condition is not satisfied in probabilistic systems. If we accept Probabilistic Effects, then the Markov Condition fails not only for controversial examples like Cartwright’s Chemical Factory case; it fails for every model which includes probabilistic effects in $\mathcal{U}$. In the causal graph shown in figure 5, for instance, the Markov Condition tells us that $X \perp \! \! \! \perp Y$. Worse still, it tells us that $Y \perp \! \! \! \perp \Pr_Y$, since $\Pr_Y \not\in \mathbf{DE}(Y)$ and $\mathbf{PA}(Y) = \emptyset$. But so long as $\Pr(Y \mid \Pr_Y = \text{pr}Y) = \text{pr}Y(Y)$, for every possible value $\text{pr}Y$, $Y$ will not be independent of $\Pr_Y$.

We can emend Probabilistic Effects by insisting that, for every $X$ in the model, $X$’s merely probabilistic causal descendants are included in $\mathbf{DE}(X)$, and that, if $Y$ is a probabilistic effect, then $\mathbf{PA}(Y) = \{\Pr_Y\}$. However, we will no longer be able to wheel in Theorem 1 to justify the probabilistic cmc, since that theorem was proven under the assumption that every endogenous variable’s value is a function of the values of its causal parents. Nor will a comparable theorem hold for these models, so emended. For the variable $\Pr_Y$ either takes as its value a joint or a marginal distribution over $Y$. If the variable $\Pr_Y$ merely provides a marginal distribution over $Y$, then there is nothing to rule out that $\Pr(\cdot)$, the objective probability function, provides a joint distribution according to which $Y \not\perp \! \! \! \perp Z \mid \Pr_Y$, for $Z \not\in \mathbf{DE}(Y)$.

If, on the other hand, the variable $\Pr_Y$ provides a joint distribution, then there’s nothing to rule out that this joint distribution is one according to which $Y \not\perp \! \! \! \perp Z \mid \Pr_Y$, for $Z \not\in \mathbf{DE}(Y)$. We could write this in by hand, but that would amount to a stipulation, and not a proof, that the probabilistic cmc holds.

The punchline: even if we ought to bring cases of probabilistic causation under the aegis of a theory of deterministic causation by appealing to the thesis of probabilistic effects, this thesis will do little to help justify the probabilistic cmc.

---

6 The objective probability function $\Pr$ must provide a joint distribution over all the variables in $\mathcal{U} \cup \mathcal{V}$, else the Markov Condition won’t even be capable of being true, since $\Pr(Y \mid \Pr_Y, Z)$ won’t be defined.
5 Modeling Probabilistic Causation

In their 1999 and 2004, Hausman & Woodward offer several arguments to the effect that the Markov Condition will hold of any deterministic Markovian causal system (at least, at some level of description or other). These arguments appeal, among other things, to the particulars of their manipulationist account of causation, codified in a condition called ‘modularity’. I accept modularity, but if all we wanted was a reason to think that the objective probability function defined over $U \cup V$ satisfies the Markov Condition in deterministic systems, these arguments can seem a bit like overkill. It seems to me that the best argument for the deterministic $cmc$ is given by Theorem 1. Given this theorem, all we need in order to argue for the deterministic $cmc$ is the assumption that, if $A$ events are caused by all and only $B$ events, then $\Pr(A) = \Pr(B)$. That is, we need only assume that the objective probability function defined on the exogenous variables induces an objective probability function on the endogenous variables in the way described by (1). This could fail if, e.g., rain was caused by all and only low atmospheric pressure, but the probability of rain was different from the probability of low atmospheric pressure. There is a position in logical space that allows this, but it is not a very attractive one. Indeed, though she rejects modularity, Cartwright appears to accept the deterministic $cmc$—she refers to it as “a well-known and trivial result for deterministic systems.”

My goal in this section is to advance a similar kind of argument for the probabilistic $cmc$—at least, given that a certain constraint is met. We saw in the previous section that the thesis of probabilistic effects is incapable of vindicating the Markov Condition except by fiat. Therefore, I will build a class of models around the thesis of probabilistic causation. Within these models, whether and how two variables are structurally related by the equations in $E$ is itself a matter of probability.

5.1 Probabilistic Causal Models

A probabilistic causal model $\mathcal{M}$ is a 4-tuple $\langle U, V, E, \rho \rangle$ of a set $U = \{U_1, U_2, ..., U_N\}$ of exogenous variables, a set $V = \{V_1, V_2, ..., V_M\}$ of endogenous variables, a set $E$ of structural equations, and a probability function $\rho$, defined over the variables in $U$ and the structural equations in $E$.

In the deterministic causal models I introduced back in section 2, for each variable $V \in V$, there was a unique structural equation, $\phi_V$, associated with that variable. In a probabilistic causal model, we allow there to be arbitrarily

---

7 Cartwright 2002, p. 424
many structural equations, $\phi_1^V, \phi_2^V, \ldots, \phi_k^V$, associated with each variable $V \in \mathcal{V}$.

$$
\mathcal{E} = \left\{ \begin{array}{cccc}
\phi_1^{V_1}, & \phi_2^{V_1}, & \ldots & \phi_k^{V_1} \\
\phi_1^{V_2}, & \phi_2^{V_2}, & \ldots & \phi_k^{V_2} \\
\vdots & \vdots & \ldots & \vdots \\
\phi_1^{V_M}, & \phi_2^{V_M}, & \ldots & \phi_k^{V_M} \\
\end{array} \right\}
$$

($k_i$ is the number of structural equations associated with the variable $V_i$.)

We introduce these additional structural equations because, when causation is probabilistic, we do not know precisely how the value of one variable is determined by the other variables in our model. $X$ could be structurally determined by $Y$ and $Z$ in the way specified by $\phi_1^X$, or it could be structurally determined by $Y$ and $W$ in the way specified by $\phi_2^X$.

For illustration, suppose that I have an probabilistic weed killer. If I spray a weed with this weed killer, there is a 90% probability that it will cause the weed to die, and a 10% probability that it will not. Hausman & Woodward (1999) advised modeling a case like this with the endogenous variables $S$, Pr$_W$, and $W$, where $S$ is a binary variable taking the value 1 if the weed killer is sprayed and taking the value 0 otherwise; $W$ is a binary variable taking the value 1 if the weed lives and 0 if it dies; and Pr$_W$ is a variable whose values are probability functions defined over the values of $W$. In contrast, I’m going to model this situation with just the binary variables $S$ and $W$, and two structural equations,

$$
\begin{align*}
\phi_1^W : (U_W, S) &\rightarrow \{0, 1\} \\
\phi_2^W : U_W &\rightarrow \{0, 1\}
\end{align*}
$$

$\phi_1^W$ says that whether the weed dies is determined in part by whether the weed killer is sprayed. If the weed killer is sprayed, then the plant dies; if the weed killer isn’t sprayed, then whether the plant lives depends upon exogenous factors (light, water, etc.). $\phi_2^W$ says that whether the weed killer is sprayed doesn’t determine whether the plant dies. Whether it lives or dies is determined solely by exogenous factors. Since it is a matter of probability whether the spraying of the weed killer will cause the weed to die, it is a matter of probability whether $S$ structurally determines $W$, as specified by the equation $\phi_1^W$, or whether it fails to structurally determine $W$, as specified by $\phi_2^W$. The function $\rho$ tells us how probable it is that $W$ is structurally determined by $S$. Since there’s a 90% probability that the weed killer, if sprayed, will kill the weed, $\rho(\phi_1^W \mid S) = 0.9$, while $\rho(\phi_2^W \mid S) = 0.1$. 

Before, we used the structural equations in $\mathcal{E}$ to induce a probability distribution over the endogenous variables via (1),

$$\mathcal{P}(V = v) = \mathcal{P}(\phi_{V}^{-1}(V = v)),$$

where $\phi_{V}^{-1}(V = v)$ is the pre-image of $V = v$ under $\phi_{V}$. Since $V$’s structural equation is no longer fixed, we cannot induce a joint distribution over the endogenous variables in this straightforward way. But there is a closely-related way to induce a probability distribution over the endogenous variables.

If a set of structural equations $\mathcal{E} \subseteq \mathcal{E}$ contains exactly one $\phi_{V_{i}}$ for every $V_{i} \in \mathcal{V}$, call that set of structural equations maximal. A maximal set of structural equations $\mathcal{E}$ specifies a unique structural equation for every endogenous variable. This, together with the variable sets, $\mathcal{U}$ and $\mathcal{V}$, and the probability distribution over the exogenous variables, constitutes one of the deterministic models familiar from §2. So we can—relative to that set of structural equations—induce a probability function over the endogenous variables according to (1). Call the resulting distribution $\mathcal{P}_{\mathcal{E}}$.

If a set of structural equations $\mathcal{E} \subseteq \mathcal{E}$ is given a non-zero probability by $\rho$, then call that set of structural equations allowed. We can induce a joint distribution over the endogenous variables of a probabilistic causal model by requiring that, for every allowed maximal subset of $\mathcal{E}$, $\mathcal{E}$, $\rho$ satisfy 2.

$$\rho(\cdot | \mathcal{E}) = \mathcal{P}_{\mathcal{E}}(\cdot)$$

Because the set of maximal subsets of $\mathcal{E}$ forms a partition, it follows that

$$\rho(\cdot) = \sum_{\mathcal{E}} \mathcal{P}_{\mathcal{E}}(\cdot)\rho(\mathcal{E}).$$

By taking each maximal, allowed subset of $\mathcal{E}$, $\mathcal{E}$, and wedding it with the original variable sets and the induced probability function $\mathcal{P}_{\mathcal{E}}$, we get a deterministic causal model, $(\mathcal{U}, \mathcal{V}, \mathcal{E}, \mathcal{P}_{\mathcal{E}})$. Call these the probabilistic causal model’s deterministic submodels. 3 tells us that the joint probability function $\rho$ defined over the endogenous variables in $\mathcal{V}$ in a probabilistic causal model is a mixture of the joint probability functions of its deterministic submodels.

$$\rho(V_{1}, V_{2}, ..., V_{M}) = \sum_{\mathcal{E}} \mathcal{P}_{\mathcal{E}}(V_{1}, V_{2}, ..., V_{M}) \cdot \rho(\mathcal{E}).$$
The Probabilistic Causal Markov Condition

The Markov Condition tells us that conditioning on the values of a variable \( X \)'s structural parents screens it off from all its non-descendants. To formulate the Markov Condition in these probabilistic models, then, we will need to say something about which variables in \( \mathcal{V} \) are \( X \)'s structural parents.

However, different structural equations may give \( X \) different causal parents. For instance, \( \mathcal{E} \) could contain both \( X := \phi^1_X(Y, U_X) \) and \( X := \phi^2_X(Z, U_X) \). In any deterministic model containing \( \phi^1_X \), \( \text{PA}(X) \) would be \( \{Y\} \). In any deterministic model containing \( \phi^2_X \), \( \text{PA}(X) \) would be \( \{Z\} \). Call \( X \)'s structural parents according to a maximal allowed subset of \( \mathcal{E} \), \( \mathcal{E}' \), \( \text{PA}_{\mathcal{E}'}(X) \).

Then, we can take \( \text{PA}(X) \) to be the union of all of the \( \text{PA}_{\mathcal{E}'}(X) \).

\[
\text{PA}(X) \overset{\text{def}}{=} \bigcup_{\mathcal{E}'} \text{PA}_{\mathcal{E}'}(X)
\]

We can similarly define the causal descendants of \( X \) to be the union of all of \( X \)'s descendants in every maximal allowed subset of \( \mathcal{E}, \mathcal{E}' \).

\[
\text{DE}(X) \overset{\text{def}}{=} \bigcup_{\mathcal{E}'} \text{DE}_{\mathcal{E}'}(X)
\]

Just as before, we will restrict our attention to those causal models which are acyclic.

**Acyclicity.** For every \( V \in \mathcal{V} \), \( \text{PA}(V) \cap \text{DE}(V) = \emptyset \).

And, just as before, we will restrict attention to those models which satisfy **Exogenous Independence**.

**Exogenous Independence.** Every exogenous variable has at most one child, and for any two disjoint sets of exogenous variables \( U, V \subseteq \mathcal{U} \),

\[
U \perp_{\rho} V
\]

As before, I’ll call any probabilistic causal model which satisfies both **Acyclicity** and **Exogenous Independence** a ‘Markovian’ model.

I show in the appendix that, given this definition of \( \text{PA}(X) \) and \( \text{DE}(X) \), the Markov Condition will hold within every deterministic submodel of a Markovian probabilistic causal model.

**Theorem 2.** For every Markovian probabilistic causal model \( \mathcal{M} = \langle \mathcal{U}, \mathcal{V}, \mathcal{E}, \rho \rangle \), every allowed maximal subset \( \mathcal{E}' \subseteq \mathcal{E} \), and every pair of sets of variables \( X, Y \subseteq \mathcal{V} \),
such that \( Y \cap \text{DE}(X) = \emptyset \),

\[
X \indep_{\mathcal{P}_E} Y \mid \text{PA}(X)
\]

(This does not follow straightaway from \textbf{Theorem 1} since \( \text{PA}_E(X) \subseteq \text{PA}(X) \) and \( \text{DE}_E(X) \subseteq \text{DE}(X) \)).

Even though the Markov Condition holds in each of these deterministic submodels, this property is not preserved under mixtures. To see why, note that

\[
\rho(X \mid Y, \text{PA}(X)) = \frac{\rho(X, Y, \text{PA}(X))}{\rho(Y, \text{PA}(X))}
\]

\[
= \frac{1}{\rho(Y, \text{PA}(X))} \cdot \sum_{\mathcal{E}} \rho(X, Y, \text{PA}(X), \mathcal{E})
\]

\[
= \frac{1}{\rho(Y, \text{PA}(X))} \cdot \sum_{\mathcal{E}} \frac{\rho(Y, \text{PA}(X), \mathcal{E})}{\rho(Y, \text{PA}(X))} \cdot \frac{\rho(X, Y, \text{PA}(X), \mathcal{E})}{\rho(Y, \text{PA}(X))}
\]

\[
= \sum_{\mathcal{E}} \frac{\rho(X, Y, \text{PA}(X), \mathcal{E})}{\rho(Y, \text{PA}(X))} \cdot \frac{\rho(Y, \text{PA}(X), \mathcal{E})}{\rho(Y, \text{PA}(X))} \cdot \rho(\mathcal{E} \mid Y, \text{PA}(X))
\]

\[
= \sum_{\mathcal{E}} \mathcal{P}_E(X \mid Y, \text{PA}(X)) \cdot \rho(\mathcal{E} \mid Y, \text{PA}(X))
\]

Therefore, there are two ways that the value of a variable \( Y \) can provide information about the value of another variable \( X \), once we’ve conditioned on the values of \( X \)'s parent variables. Firstly, it could provide information about the value of \( X \) within a submodel—if, e.g., \( \mathcal{P}_E(X \mid Y, \text{PA}(X)) \neq \mathcal{P}_E(X \mid \text{PA}(X)) \). Secondly, it could provide information about which submodel has actually obtained—if, e.g., \( \rho(\mathcal{E} \mid Y, \text{PA}(X)) \neq \rho(\mathcal{E} \mid \text{PA}(X)) \). Even though \textbf{Theorem 2} assures us that conditioning on the values of \( X \)'s parent variables will prevent \( Y \) from providing information in the first way, conditioning on the values of \( X \)'s parents need not, and in general will not, prevent \( Y \) from providing information in the second way.

For an example in which the values of \( X \)'s parent do not prevent \( Y \) from providing information in this second way, consider the probabilistic causal model \( \mathcal{M}_2 = \langle U_2, V_2, E_2, \rho_2 \rangle \), shown in figure 7. There, \( F, C, \) and \( E \) are each binary variables. \( \phi_F, \phi_C^1, \phi_C^2, \phi_E^1, \) and \( \phi_E^2 \) are all just the identity function. \( \rho_2 \) tells us that there is an 80% probability that the values of \( C \) and \( E \) will be the same as \( F \) (as specified by the functions \( \phi_C^1 \) and \( \phi_E^1 \)) and a 20% probability
\[ V_2 = \{F, C, E\} \quad U_2 = \{U_F, U_C, U_E\} \]

\[ \mathcal{E}_2 = \left\{ \begin{array}{l}
F := \phi_F(U_F) \\
C := \phi_C^1(F) \\
E := \phi_E^1(F) \\
\end{array} \right\} \]

\[ \mathcal{E}_2 = \left\{ \begin{array}{l}
C := \phi_C^2(U_C) \\
E := \phi_E^2(U_E) \\
\end{array} \right\} \]

\[ \rho_2(\phi_F) = 1 \]

\[ \rho_2(\phi_C^1) = \rho_2(\phi_E^1) = 0.8 \]

\[ \rho_2(\phi_C^2) = \rho_2(\phi_E^2) = 0.2 \]

\[ \rho_2(U_F) = \rho_2(U_C) = \rho_2(U_E) = 1 \]

\[ \phi_F, \phi_C^i, \phi_E^i : x \mapsto x, \ i = 1, 2 \]

\[ \mathcal{E}_1 = \{\phi_F, \phi_C^1, \phi_E^1\} \]

\[ \mathcal{E}_2 = \{\phi_F, \phi_C^2, \phi_E^2\} \]

And \( \rho_2 \) tells us that the structural determination relations \( \mathcal{E}_1 \) have an 80% probability of obtaining, while the structural determination relations \( \mathcal{E}_2 \) have a 20% probability of obtaining.

Even though each of the deterministic submodels \( \langle U_2, V_2, \mathcal{E}_1^1, \rho_{\mathcal{E}_1^1} \rangle \) and \( \langle U_2, V_2, \mathcal{E}_2^2, \rho_{\mathcal{E}_2^2} \rangle \) satisfy the Markov Condition, their mixture will not. For

\[ \rho_2(C \mid F) = \sum_{\mathcal{E}} \mathcal{P}_{\mathcal{E}}(C \mid F) \cdot \rho_2(\mathcal{E} \mid F) \]

\[ = \mathcal{P}_{\mathcal{E}_1}(C \mid F) \cdot \rho_2(\mathcal{E}_1 \mid F) + \mathcal{P}_{\mathcal{E}_2}(C \mid F) \cdot \rho_2(\mathcal{E}_2 \mid F) \]

\[ = 1 \cdot 0.8 + 0 \cdot 0.2 \]

\[ = 0.8 \]

while

\[ \rho_2(C \mid F, E) = \sum_{\mathcal{E}} \mathcal{P}_{\mathcal{E}}(C \mid F, E) \cdot \rho_2(\mathcal{E} \mid F, E) \]

\[ = \mathcal{P}_{\mathcal{E}_1}(C \mid F, E) \cdot \rho_2(\mathcal{E}_1 \mid F, E) + \mathcal{P}_{\mathcal{E}_2}(C \mid F, E) \cdot \rho_2(\mathcal{E}_2 \mid F, E) \]
\[ 1 \cdot 1 + 0 \cdot 0 = 1 \]

The value of \( E \) gives us extra information about the value \( C \) over and above the information given by the value of \( F \). Knowing the value of \( E \) tells us whether \( \mathcal{E}_1 \) or \( \mathcal{E}_2 \) actually obtains. That is, knowing the value of \( E \) tells us whether or not \( E \) is structurally determined by \( F \) or not. And this tells us whether \( C \) is structurally determined by \( F \) or not. But then \( C \not\perp_{p_2} E \mid F \), in violation of the Markov Condition.

When causation goes probabilistic, we get violations of the Markov Condition. The probabilistic causal models show us precisely how this occurs. A variable \( X \)'s non-descendants can tell us something about which structural determination relations actually obtain, which can tell us something about the value of a variable \( X \), even once we've conditioned on the values of \( X \)'s structural parents.

Just because we find some violations of the Markov Condition, this does not spell disaster for the statistical techniques of Pearl and Spirtes et al. when the systems under study are probabilistic—for we do not yet know how deep the violations cut. We should remind ourselves that almost every statistical technique utilized by practicing researchers require that the system under study meet certain non-trivial conditions. Verifying that these conditions are met is often a difficult task (perhaps prohibitively so, requiring the researcher to make bold assumptions and hope that the world plays along). Having a certain class of exceptions is by no means a fatal flaw for a statistical method; we should, however, be up front and honest about these exceptions.

Correlatively, once counterexamples to the Markov Condition are identified, the proper response is not to chalk up these violations to probabilistic causation and abandon all statistical methods predicated on the \( \text{cmc} \) whenever we are dealing with a probabilistic system. Rather, we ought to shift our attention to the following question: under what general conditions can we expect the Markov Condition to hold, even in probabilistic contexts? Here, the news is relatively good. There is a general condition which it seems reasonable to expect a wide variety of probabilistic systems to obey which is, all by itself, sufficient to guarantee that the Markov Condition is satisfied. I call this condition ‘structural independence’.

**Structural Independence.** For all \( X, Y \subseteq \mathcal{V} \) such that \( Y \cap \text{DE}(X) = \emptyset \),

\[
\phi_X \perp Y, \phi_Y \mid \text{PA}(X)
\]
where \( \phi_V \overset{\text{def}}{=} \bigcup_V \phi_V \), for all \( V \in V \).

**Structural independence** is satisfied when which structural equations govern the \( X \in X \) is independent of both the values of the \( Y \in Y \) and which structural equation govern the \( Y \in Y \), for any \( X, Y \) such that \( Y \cap \text{DE}(X) = \emptyset \). This amounts to a kind of locality constraint. It says that the way that \( X \) is determined by its structural parents is only affected by the values of those parents, and not by the values of the other variables upstream of it, nor which structural determination relations upstream of it actually obtain.

In Appendix A.2, I offer a proof of the following theorem:

**Theorem 3.** If structural independence holds, then, for all \( X, Y \subseteq V \) such that \( Y \cap \text{DE}(X) = \emptyset \),

\[ X \perp Y \mid \text{PA}(X) \]

That is, if \( \rho \) tells us that the probability that a set of variables are structurally determined by their parents in any particular way is unaffected by the values of the variables or the structural determination relations which obtain upstream of them, then the Markov Condition will be satisfied. (The converse does not hold, by the way; there are systems which violate structural independence but which nevertheless satisfy the Markov Condition.) This is just a formal result about the probabilistic causal models I’ve introduced here. In the next subsection, I’ll explain what I take the philosophical significance of this result to be.

### 5.3 Structural Independence and Autonomy

I think we can begin to make sense of this result by looking at Hausman & Woodward (1999)’s other response to Cartwright (the one I put off back in section 3). Here’s what Hausman & Woodward say: if we are to make sense of the situation described in *Chemical Factory*, where a correlation between variables lingers even after we have conditioned on all the causal parents in the graph, then there must be some other variable which is causally intermediate between \( F \) and \( C \) and \( E \), and which explains any correlations between \( C \) and \( E \) which remain after conditioning on \( F \). It is after all a familiar fact, they remind us, that in a graph like that shown in figure 8, even though conditioning on \( W \) will render \( Y \) and \( Z \) probabilistically independent, conditioning on \( X \) need not. (This is because \( W \)’s exogenous parent, \( U_W \), will introduce correlations between \( Y \) and \( Z \) which will remain even after conditioning on \( X \).)

---

8 Similar remarks can be found in Pearl (2000)
So, if there is a variable intermediate between $F$ and $C$ and $E$—call it ‘$W$’—then we should expect that $C$ and $E$ remain correlated even after we condition on the value of $F$. Moreover, Hausman & Woodward assert, there surely is such a variable between $F$ and $C$ and $E$—it is the variable telling us the state of the factory, whether chemical production was initiated or not. In that case, the correct causal graph is just the one given by the solid black arrows in figure 9. The original arrows, between $F$ and $C$ and $F$ and $E$, were not, therefore, correct; they told us (accurately) that $F$ was causally relevant to $C$ and $E$, but they also told us (inaccurately) that the way that $C$ is structurally determined by $F$ is autonomous of the way that $E$ is structurally determined by $F$. What we have here is a failure of causal sufficiency. A variable set $\mathcal{V}$ is said to be causally sufficient just in case no common cause of any of the variables $\mathcal{V}$ is excluded from $\mathcal{V}$. However, the Markov Condition, Hausman & Woodward (1999) and Spirtes et al. (2000) claim, should be restricted to only those variable sets which are causally sufficient.

We can use this diagnosis of what’s going on in Chemical Factory to see exactly how it could happen that whether/how a cause $(C)$ acts to bring about one of its effects ($E_1$) is not probabilistically independent of whether/how it acts to bring about another of its effects ($E_2$). That is, this diagnoses how structural independence could be violated by a causally probabilistic system. This can happen when the structural determination relations between $C$ and $E_1$ and between $C$ and $E_2$ are not autonomous—where I say that a structural determination relation is autonomous in a causal model $\mathcal{M}$ iff the following condition is met.

**Autonomy:** In a causal graph $\mathcal{G}$ over the variables in $\mathcal{V}$, the structural determination relation represented by $Y \rightarrow X$ is autonomous of the structural determination relation represented by $W \rightarrow X$ iff there is no correct causal graph $\mathcal{G}'$ over the variables in $\mathcal{V}'$ such that $\mathcal{V} \subseteq \mathcal{V}'$, and in $\mathcal{G}'$, $Y, W \in \text{AN}(O)$ and $O \in \text{AN}(X)$.
Figure 10: Assuming that a correct causal graph for a certain variable set is given by the black arrows, the structural determination relations associated with the arrows in the graph for the restricted variable set (given by the greyscale arrows) will not all be autonomous. 

(‘AN(X)’ is a vector of X’s ancestors.) That is: if there’s a way of tying together two of the arrows leading from PA(X) to X, so that some of X’s parents only influence X via some common intermediary omitted variable O, then these arrows do not represent autonomous structural determination relations. Autonomy is the endogenous component of causal sufficiency. Assuming that there are no omitted common causes of any of the exogenous variables in the graph, autonomy is sufficient for causal sufficiency. And assuming that Exogenous Independence is satisfied, the only way that a failure of causal sufficiency could lead to violations of the Markov condition is via a failure of autonomy. We needn’t have a fork in order to have structural determination relations which are not autonomous. Consider a causal graph in which $X \rightarrow Y \rightarrow Z$ and $X \rightarrow Z$, where we could enrich the variable set in such a way as to recover the causal graph in figure 10(b) (the one given by the black arrows). In this case, the structural determination relation $X \rightarrow Z$ is not autonomous of the structural determination relation $Y \rightarrow Z$. This is, presumably, what happens in the case of exposure to disease and the presence of symptoms—the structural determination relation between one symptom and exposure is not autonomous of the structural determination relation between another symptom and exposure. For we could include a variable describing whether I have contracted the disease, and then the exposure would only influence the symptoms through this intermediary omitted variable.

One lesson to take from this—the lesson which Hausman & Woodward and Spirtes et al. take from this—is just that we ought not use statistical methods which rest on the truth of the cmc when we have reason to think that the variable set under investigation isn’t causally sufficient, and therefore that the structural determination relations in the graph are not autonomous. And, to a good extent, the probabilistic models introduced here justify this reaction. They show us that, when we think that we are dealing with a system
which violates *structural independence*, we ought not expect it to conform to the cmc.

If we additionally suppose that every causal system which violates *structural independence* does so because some of the structural determination relations it represents are not autonomous, and therefore, the variable set we are looking at is causally insufficient, then we will think, like Hausman & Woodward, that all failures of the Markov Condition are attributable to failures of causal sufficiency. However, we might be a bit hesitant to make this final supposition. After all, there are Quantum Mechanical systems which have the same formal structure as Chemical Factory, but for which there are no hidden variables which could be included in the variable set. We do not yet have any reason to think that, in general, whenever structural independence is violated, this will be attributable to a violation of autonomy.

I won’t take a stand on this question here. I am content to have isolated and clarified the bone of contention between Cartwright and her dialectical opponents. On my understanding, that disagreement boils down to differing answers to the following question: are there causally probabilistic systems (outside of the realm of Quantum Mechanics) which violate *structural independence* without violating autonomy? Hausman & Woodward, Spirtes et al., and Pearl answer ‘no.’ Cartwright answers ‘yes.’

6 In Summation

I take the central conclusions of this essay to be the following: firstly, even if the thesis of probabilistic effects is true, we cannot use it to vindicate the cmc. Secondly, the thesis of probabilistic causation fails to validate the Markov Condition. However, the probabilistic causal systems which violate the Markov Condition are restricted to those which also violate *structural independence*. Finally, we ought to understand the dispute about whether or not probabilistic causal systems conform to the Markov Condition as a dispute about whether or not any autonomous causal structures violate structural independence.

9 To be a bit more careful: no *local* hidden variable theory is capable of predicting the observed violations of the Bell inequalities; however, if we accept Bohmian mechanics (a non-local, deterministic version of QM), then there will be variables which we could include in the variable set which would restore the Markov Condition.
A Appendices

A.1 Modeling Probabilistic Causation with Dummy Variables

Steel (2005) and Glymour (2010) propose modelling probabilistic causation with dummy variables. To illustrate, suppose that we have a system in which a binary variable \( X \) indeterministically influences a binary variable \( Y \). Suppose that there is an 80\% chance that the value of \( Y \) will equal the value of \( X \) and a 20\% chance that \( Y \) will be 0. Steel and Glymour model situations like these by including the dummy variable \( D_Y \), which is treated formally as an exogenous parent of \( Y \), and writing \( Y \)'s structural equation as follows:

\[
Y := D_Y \cdot X
\]

If we then let \( P(D_Y = 1) = 0.8 \) and \( P(D_Y = 0) = 0.2 \), we will get a model in which there is an 80\% chance that \( Y \) will take on the same value as \( X \) and a 20\% chance that \( Y \) will be 0, as we wanted.

This approach to modeling probabilistic causation is just as expressive as my own. If a variable \( V \in \mathcal{V} \) has \( N \) structural equations \( \phi^1_V, \ldots, \phi^N_V \in \mathcal{E} \), then we can simply use a dummy variable \( D_{\phi_V} \) with \( N \) values, and let

\[
V := \Phi_V (D_{\phi_V}, \text{PA}(V), U_V)
\]

where \( \Phi_V \) is a function from values of \( D_{\phi_V} \) to functions from \( \langle \text{PA}(V), U_V \rangle \) to \( V \).

Within this framework, Steel argues that it follows from Exogenous Independence and Theorem 1 that the Markov condition will be satisfied even in cases of probabilistic causation. Drouet (2009) responds that, when we are dealing with dummy variables, rather than genuine variables which represent exogenous causes, there is no justification for imposing Exogenous Independence. Another worry about this approach to vindicating the probabilistic Markov condition is that, in the presence of Exogenous Independence, the dummy variable approach does not allow the value that a variable takes on to affect the probability that it influences its descendants in some particular way. For instance, suppose that we have two binary variables, \( X \) and \( Y \), with \( \text{PA}(Y) = \{X\} \), and that if \( X = 1 \), then there is an 80\% chance that the value of \( Y \) will be 1 and a 20\% chance that it will be 0, whereas, if \( X = 0 \), then there is an 70\% chance that the value of \( Y \) will be 1 and a 30\% chance that it will be 0. Within the probabilistic causal models introduced in §5.1, this is accomplished straightforwardly by having two structural equations, \( \phi^1_Y \) and \( \phi^2_Y \), where \( \phi^1_Y \) determines \( Y \)'s value to be the same as \( X \)'s, and
\( \phi \) determines \( Y \)'s value to be different from \( X \)'s. Then, since the way that \( Y \)'s value is determined depends upon which value \( X \) takes on,

\[
\rho(\phi_1 | X = 1) = 0.8 \neq 0.3 = \rho(\phi_1 | X = 0)
\]

Within the models from §5.1, we may have a variable \( V \)'s structural determination relation be probabilistically dependent upon the values of \( \text{PA}(V) \). Moreover, this is entirely consistent with structural independence, and therefore entirely consistent with the satisfaction of the Markov condition. However, if we attempted to model this situation with a dummy variable \( D_Y \), then the value of \( D_Y \) would be probabilistically dependent upon the value of \( X \), which would entail a violation of Exogenous Independence. Thus, Theorem 1 could not be used to guarantee that the Markov condition would be satisfied. So while the dummy variable approach to modeling probabilistic causation is just as expressive as the approach offered here, the vindication of the Markov condition that goes by way of dummy variables, Exogenous Independence, and Theorem 1 is not as extensive as the vindication that goes by way of structural independence.

### A.2 Proofs

**Theorem 2.** For every Markovian probabilistic causal model \( M = \langle V, U, E, \rho \rangle \), every allowed maximal subset \( \mathcal{E} \subseteq E \), and every pair of sets of variables \( X, Y \subseteq V \) such that \( Y \cap \text{DE}(X) = \emptyset \),

\[
X \perp_{\mathcal{E}} Y \mid \text{PA}(X)
\]

where

\[
\text{DE}(X) = \bigcup_{\mathcal{E}} \text{DE}_{\mathcal{E}}(X) \quad \text{and} \quad \text{PA}(X) = \bigcup_{\mathcal{E}} \text{PA}_{\mathcal{E}}(X)
\]

*Proof.*

\[
\text{PA}(X) \cap \text{DE}(X) = \emptyset
\]

by Acyclicity. Thus,

\[
(\text{PA}(X) - \text{PA}_{\mathcal{E}}(X)) \cap \text{DE}(X) = \emptyset
\]

And since \( \text{DE}_{\mathcal{E}}(X) \subseteq \text{DE}(X) \),

\[
(\text{PA}(X) - \text{PA}_{\mathcal{E}}(X)) \cap \text{DE}_{\mathcal{E}}(X) = \emptyset \quad \text{and} \quad Y \cap \text{DE}_{\mathcal{E}}(X) = \emptyset
\]

Since \( \langle V, U, \mathcal{E}, \mathcal{P} \rangle \) is just a Markovian causal model, it follows from Theorem 1 that

\[
X \perp_{\mathcal{P}} \text{PA}(X) - \text{PA}_{\mathcal{E}}(X), Y \mid \text{PA}_{\mathcal{E}}(X)
\]
Since $A \perp B, C \mid D \models A \perp B \mid C, D,$
\[
X \perp_{\mathcal{E}} Y \mid \text{PA}(X) - \text{PA}_{\mathcal{E}}(X), \text{PA}_{\mathcal{E}}(X)
\]
\[
X \perp_{\mathcal{E}} Y \mid \text{PA}(X)
\]

\[\Box\]

**Lemma 1.** For all $Y \subseteq \mathcal{V}$ such that $Y \cap \text{DE}(X) = \emptyset,$
\[
\rho(X \mid \text{PA}(X), Y, \phi_X) = \rho(X \mid \text{PA}(X), \phi_X)
\]

**Proof.**
\begin{align*}
(1) \quad & \rho(X \mid \text{PA}(X), Y, \phi_X) = \sum_{\mathcal{E} \supseteq \phi_X} \rho(X \mid \text{PA}(X), Y, \mathcal{E}) \\
(2) \quad & = \sum_{\mathcal{E} \supseteq \phi_X} \rho(X \mid \text{PA}(X), \mathcal{E}) \\
(3) \quad & = \rho(X \mid \text{PA}(X), \phi_X)
\end{align*}

Line 2 follows from line 1 because each submodel $\mathcal{E}$ is a deterministic Markovian model, and so it satisfies the Markov Condition. \[\Box\]

**Lemma 2.** Given *Structural Independence*,
\[
\rho(\phi_X \mid Y, \text{PA}(X)) = \rho(\phi_X \mid \text{PA}(X))
\]

**Proof.** From *Structural Independence*, $\rho(\phi_X \mid Y, \phi_Y, \text{PA}(X)) = \rho(\phi_X \mid \text{PA}(X))$. Then,
\[
\rho(\phi_X \mid Y, \text{PA}(X)) = \sum_{\phi_Y} \rho(\phi_X, \phi_Y \mid Y, \text{PA}(X)) \\
= \sum_{\phi_Y} \rho(\phi_X \mid \phi_Y, Y, \text{PA}(X)) \cdot \rho(\phi_Y \mid Y, \text{PA}(X)) \\
= \rho(\phi_X \mid \text{PA}(X)) \cdot \sum_{\phi_Y} \rho(\phi_Y \mid Y, \text{PA}(X)) \\
= \rho(\phi_X \mid \text{PA}(X))
\]

\[\Box\]

**Theorem 3.** If, for all $X, Y \subseteq \mathcal{V}$ such that $Y \cap \text{DE}(X) = \emptyset,$
\[
\phi_X \perp Y, \phi_Y \mid \text{PA}(X)
\]
then, for all $X, Y \subseteq \mathcal{V}$ such that $Y \cap \text{DE}(X) = \emptyset,$
\[
X \perp Y \mid \text{PA}(X)
\]
Proof.

(1) \( \rho(X \mid Y, PA(X)) = \sum_{\phi_X} \rho(X, \phi_X \mid Y, PA(X)) \)

(2) \( = \sum_{\phi_X} \rho(X \mid \phi_X, Y, PA(X)) \cdot \rho(\phi_X \mid Y, PA(X)) \)

(3) \( = \sum_{\phi_X} \rho(X \mid \phi_X, Y, PA(X)) \cdot \rho(\phi_X \mid PA(X)) \)

(4) \( = \sum_{\phi_X} \rho(X \mid \phi_X, PA(X)) \cdot \rho(\phi_X \mid PA(X)) \)

(5) \( = \sum_{\phi_X} \rho(X, \phi_X \mid PA(X)) \)

(6) \( = \rho(X \mid PA(X)) \)

Line 3 follows from line 2 by lemma 2, and line 4 follows from line 3 by lemma 1.

References


