# PROBABILISTIC ACTUAL CAUSATION

LUKE FENTON-GLYNN

DEPARTMENT OF PHILOSOPHY, UNIVERSITY COLLEGE LONDON

GOWER STREET, LONDON, WC1E 6BT, U.K.

ABSTRACT. Actual (token) causes – e.g. Suzy's being exposed to asbestos – often bring about their effects – e.g. Suzy's suffering mesothelioma – probabilistically. I use probabilistic causal models to tackle one of the thornier difficulties for traditional accounts of probabilistic actual causation: namely probabilistic preemption.

Luke Fenton-Glynn

## 1. Introduction

Actual (token) causation is the relation that obtains when, for example, Suzy's being exposed to asbestos causes her to suffer mesothelioma. A number of theorists (e.g. Halpern and Pearl 2001, 2005; Hitchcock 2001, 2007; Weslake 2016) have deployed structural equations models (SEMs) in developing novel solutions to difficulties confronting traditional accounts of this relation. These theorists have focused on *deterministic* actual causation (DAC).[1] I draw on probabilistic causal models (PCMs) – analogues of deterministic SEMs – to provide an account of probabilistic actual causation (PAC). I don't attempt to show that my account can handle the full battery of test cases discussed in the literature. I simply demonstrate that it yields an elegant treatment of one very central case – probabilistic preemption – with a view to motivating further investigation of formal approaches to PAC.

## 2. Probability-Raising

*Probability-raising* is central to the account developed here – as on traditional accounts of PAC.[2] To explain how I will understand that notion a bit of stage-setting is required.

I take the relata of the actual causal relation to be variable values. Adopting Goldszmidt and Pearl's (1992, 669–70) notation, $P(W = w|do(V = v))$ represents the probability for $W = w$ that *would* obtain if $V$ were set to $V = v$ by an 'intervention' (Woodward 2005, 98). This is liable to diverge from the conditional probability $P(W = w|V = v)$: witness the difference between the probability of a storm *conditional* upon the barometer needle pointing toward the

[1] Cf. Halpern and Pearl (2005, 852); Hitchcock (2007, 498).

[2] Reichenbach (1971, 204); Suppes (1970); Lewis (1986, 175–84); Menzies (1989). The deficiencies of these accounts have been demonstrated by e.g. Salmon (1984, 192–202); Menzies (1996, 85–96); Hitchcock (2004).

word ‘storm’ and the probability of a storm if I had intervened upon the barometer needle to point it toward ‘storm’.

Variable $X$ taking value $X = x$ (rather than $X = x'$) raises the probability of $Y = y$ in the relevant sense iff:[3]

$$P(Y = y|do(X = x)) > P(Y = y|do(X = x')) \tag{1}$$

Appealing to interventionist probabilities means avoiding probability-raising relations between independent effects of a common cause, such as the barometer reading and the storm (cf. Lewis 1986, 178).

Probabilistic preemption cases illustrate that straightforward probability-raising is neither necessary nor sufficient for causation (Menzies 1989, 1996).

## 3. Probabilistic Preemption

The following example is inspired by Anscombe (1971).[4]

[3]Here and throughout, the probabilities (chances) should be taken to be those obtaining immediately after the interventions bringing about the variable values specified in the scope of the $do(\cdot)$ function have occurred (cf. Lewis 1986, 177).

[4]The probabilities involved (except the decision probabilities) are quantum and therefore objective and able underwrite causal relations. (If you’re worried that the decision probabilities are not objective, the example could be complicated so that the decisions are made on the basis of outcomes of quantum measurements.) I find it plausible that the probabilities of many high level sciences are also objective (cf. e.g. Loewer 2001; Ismael 2009).

**(ProbPre)** *Someone (neither you nor I) has connected a Geiger counter to a bomb so that the bomb will explode if the Geiger registers above a threshold reading. I place a place a chunk of U-232 (half-life = 68.9 years; decays by $\alpha$-emission) near the Geiger. By chance, enough U-232 atoms decay within a short enough interval for the Geiger to reach the threshold reading so that the bomb explodes. Unbeknownst to me, you've been standing nearby observing. You have a chunk of Th-228 (half-life = 1.9 years; decays by $\alpha$-emission), which contains many more atoms than my chunk of U-232. You've decided that you'll place your Th-228 near the Geiger iff I fail to place my U-232 near the Geiger. There's a negligible chance that you won't follow the course of action you've decided on. Seeing that I place my U-232 near the Geiger, you don't place your Th-228 near the Geiger.*[5]

Let $M$, $D$, $Y$, $T$, and $E$ be binary variables which, respectively, take value 1 if the following things occur (and 0 otherwise): I place my U-232 near the Geiger; you decide to place your Th-228 near the Geiger iff I don't place my U-232 near the Geiger; you place your Th-228 near the Geiger; the threshold reading is reached; the bomb explodes.

My act ($M = 1$) was an actual cause of the explosion ($E = 1$). Yet plausibly the following inequality holds:

(2) $$P(E = 1|do(M = 1)) < P(E = 1|do(M = 0))$$

[5]The range of $\alpha$-particles is 3-5 cm. Suppose that, for each of us, a decision to place our chunk 'near' the Geiger counter is a decision to place it $<$ 5cm away and a decision not to place it nearby is a decision to place it nowhere near ($\gg$ 5cm away).

That is, my placing my U-232 near the Geiger *lowers* the probability of the bomb exploding because it strongly lowers the probability of your placing your more potent Th-228 near the Geiger. Probability-raising is therefore unnecessary for actual causation.

Your decision ($D = 1$) was *not* an actual cause of the explosion, since you don't place your Th-228 near the Geiger. Yet provided there's some chance that $M = 0$, the following inequality holds:

(3) $$P(E=1|do(D=1)) > P(E=1|do(D=0))$$

Inequality (3) holds because your decision raises the probability that the bomb will still explode in the scenario in which $M = 0$.[6] Probability-raising is therefore insufficient for actual causation.

Actual causation therefore can't be identified with probability-raising. In developing a more nuanced analysis, it is helpful to appeal to PCMs.

## 4. PCMs

A PCM, $\mathscr{M}$, is a 5-tuple $\langle \mathscr{V}, \mathscr{C}, \Omega, \mathscr{F}, do(\cdot) \rangle$. $\mathscr{V}$ is a set of variables. Suppose $\mathscr{R}$ denotes a function from elements of $\mathscr{V}$ to sets of values: for all $V \in \mathscr{V}$, $\mathscr{R}(V)$ is the *range* of $V$. In Halpern and Pearl's (2005, 851–2) terminology, a formula $V_i = v_i$, for $V_i \in \mathscr{V}$ and $v_i \in \mathscr{R}(V)$, is a *primitive event*. $\mathscr{C}$ is the set of all those possible conjunctions of primitive

[6] $D = 0$ is multiply realizable: there is more than one alternative to the decision that you in fact make. E.g. you could decide that you will place your Th-228 near the Geiger no matter what, or that you will not do so no matter what. We can stipulate that the latter alternative is much more probable.

events, $V_1 = v_1 \& \ldots \& V_n = v_n$, such that $V_i \in \mathscr{V}$ and $v_i \in \mathscr{R}(V_i)$ and such that, for no pair of conjuncts $V_i = v_i$, $V_j = v_j$ is $V_i \equiv V_j$, and where no two elements of $\mathscr{C}$ differ *only* in the permutation of their conjuncts. Such a conjunction is denoted $\mathbf{V} = \mathbf{v}$ (primitive events and the null event are limiting cases of such conjunctions). Abusing notation, the fact that $v_i \in \mathscr{R}(V_i)$ for each primitive event $V_i = v_i$ in the conjunction $\mathbf{V} = \mathbf{v}$, is abbreviated $\mathbf{v} \in \mathscr{R}(\mathbf{V})$ and the set of variables that appear in $\mathbf{V} = \mathbf{v}$ is denoted $\mathbf{V}$.

Call a conjunction $\mathbf{V} = \mathbf{v}$ *maximal* if it contains a conjunct of the form $V_i = v_i$ for each $V_i \in \mathscr{V}$. $\Omega$ is the set of all maximal conjunctions of primitive events. $\mathscr{F}$ is a sigma algebra on $\Omega$. Finally, $do(\cdot)$ is a function from elements of $\mathscr{C}$ to probability distributions on $\mathscr{F}$ (cf. Pearl 2009, 70, 110): for each element $\mathbf{V} = \mathbf{v}$ of $\mathscr{C}$, $P(\cdot|do(\mathbf{V} = \mathbf{v}))$ is the probability (chance) distribution on $\mathscr{F}$ that *would* obtain if interventions were performed to bring about $\mathbf{V} = \mathbf{v}$.

A PCM can be represented graphically by taking the variables in $\mathscr{V}$ as nodes and drawing a directed edge from $V_i$ to $V_j$ ($V_i, V_j \in \mathscr{V}$) iff, where $\mathbf{S} = \mathscr{V} \backslash V_i, V_j$, there is some assignment of values $\mathbf{s}' \in \mathscr{R}(\mathbf{S})$, some pair of values $v_i, v_i' \in \mathscr{R}(V_i)$ ($v_i \neq v_i'$) and some value $v_j \in \mathscr{R}(V_j)$ such that $P(V_j = v_j | do(V_i = v_i \& \mathbf{S} = \mathbf{s}')) \neq P(V_j = v_j | do(V_i = v_i' \& \mathbf{S} = \mathbf{s}'))$.

In constructing a PCM, $\mathscr{M}_{Pre}$, of **(ProbPre)** we might take the variable set to be $\mathscr{V}_{Pre} = \{D, M, Y, T, E\}$. The range of each variable in $\mathscr{V}_{Pre}$ is the pair $\{0, 1\}$. $\mathscr{C}_{Pre}$, $\Omega_{Pre}$, and $\mathscr{F}_{Pre}$ are generated by $\mathscr{V}_{Pre}$ and $\mathscr{R}_{Pre}$ in the way described above. For each element of $\mathscr{C}_{Pre}$, the function $do(\cdot)$ returns the chance distribution on $\mathscr{F}_{Pre}$ that would obtain if interventions were performed to bring about that element of $\mathscr{C}_{Pre}$. The graph for $\mathscr{M}_{Pre}$ is given as figure 1.

$$\begin{tikzcd} D \arrow[r] & Y \arrow[dr] & & \\ M \arrow[ur] \arrow[rr] & & T \arrow[r] & E \end{tikzcd}$$

FIGURE 1

A directed path in a graph is an ordered sequence of nodes, $\langle V_1, V_2, \ldots, V_n \rangle$, such that there is a directed edge from $V_1$ to $V_2$, and a directed edge from $V_2$ to $\ldots V_n$. $\langle M, Y, T, E \rangle$ is an example of a directed path in the graph of $\mathscr{M}_{Pre}$.

## 5. Appropriate Models

In Section 6, I provide a definition of what it is for $X = x$ (rather than $X = x'$) to count as an actual cause of $Y = y$ *relative to a PCM*. I then define a non-model-relativized notion of actual causation by saying that $X = x$ (rather than $X = x'$) counts as an actual cause of $Y = y$ *simpliciter* provided that $X = x$ (rather than $X = x'$) counts as an actual cause $Y = y$ relative to at least one *appropriate* PCM.[7] A similar strategy is commonly adopted by those analyzing DAC in terms of SEMs (Hitchcock 2001, 287, 2007, 503; Weslake 2016). This requires an account of 'appropriate' models.

Many of the criteria for an appropriate SEM for evaluating DAC carry over to PCMs, including the following three:

(Partition) For all $V \in \mathscr{V}$, the elements of $\mathscr{R}(V)$ should form a partition (Halpern and Hitchcock 2010, 397–8; Blanchard and Schaffer 2016)

(Independence) For no two variables $V, W \in \mathscr{V}$ should there be elements $v \in \mathscr{R}(V)$ and $w \in \mathscr{R}(W)$ such that the states of affairs represented by $V = v$ and $W = w$ are logically or metaphysically related (Hitchcock 2001, 287; Halpern and Hitchcock 2010, 397)

[7] As the parentheses indicate I define a *contrastive* relation of actual causation. Where variables are binary – as in $\mathscr{M}_{Pre}$ – this is inconsequential and I will typically suppress such parentheses. But it becomes important in cases of multi-valued variables (see Halpern and Pearl 2005, 859).

(Naturalness) For all $V \in \mathscr{V}$, $\mathscr{R}(V)$ should include only values that represent reasonably natural and intrinsic states of affairs. (Blanchard and Schaffer 2016)

The analysis of actual causation proposed below takes all and only values of distinct variables to be potential causal relata. (Partition) insures that we don't thereby miss actual causal relations because they obtain between the values of a single variable. (Independence) insures that we don't mistake stronger-than-causal relations for causal relations. (Naturalness) insures that unnatural or non-intrinsic states of affairs do not get counted as causes and effects (see Lewis 1986, 190, 263; Paul 2000, 245).[8]

A further condition is that a model is appropriate for evaluating whether $X = x$ is an actual cause of $Y = y$ in world $\theta$ only if it satisfies (Veridicality):

(Veridicality) For any conjunction $\mathbf{V} = \mathbf{v} \in \mathscr{C}$ taken as an input, the probability distribution $P(\cdot|do(\mathbf{V} = \mathbf{v}))$ yielded as an output by $do(\cdot)$ should be the *objective chance* distribution over $\mathscr{F}$ that would$_\theta$ result from interventions setting $\mathbf{V} = \mathbf{v}$. ('Would$_\theta$' indicates that what is required is that this counterfactual be true in $\theta$.)

(Veridicality) is an analogue – for PCMs – of the requirement that SEMs encode only true counterfactuals (Hitchcock 2001, 287, 2007, 503).

In the DAC/SEMs literature another condition on model appropriateness is typically added:

(Serious Possibilities) $\mathscr{V}$ should not be such as to generate elements of $\Omega$ that represent possibilities "that we consider to be too remote" (Hitchcock 2001, 287;

[8]If *absences* are unnatural states of affairs (cf. Lewis 1986, 189–93), we might instead require that each variable have *at most one value* representing such a state of affairs.

cf. Woodward 2005, 86–91, Weslake 2016, Blanchard and Schaffer 2016).

We likely need this requirement too. A discussion of whether the vagueness and subjectivity thereby introduced is problematic would take us too far afield.[9] Still, it doesn't put the present account in any *worse* shape than its deterministic analogues. Moreover, traditional accounts of actual causation – which don't appeal to causal models – also stand in need of appeal to 'serious possibilities' (Woodward 2005, 86–8).

A final requirement – similar to one imposed in the DAC/SEM literature – for a model $\mathscr{M}$ to be an appropriate one for evaluating whether $X = x$ is an actual cause of $Y = y$ in world $\theta$ is:

(Stability) There is no model $\mathscr{M}^*$ (satisfying Partition, Independence, Naturalness, Veridicality, and Serious Possibilities) with a variable set $\mathscr{V}^*$ such that $\mathscr{V}^* \supset \mathscr{V}$ relative to which $X = x$ (rather than $X = x'$) is *not* an actual cause of $Y = y$. (Halpern and Hitchcock 2010, 394–5; Blanchard and Schaffer 2016; Halpern 2014; Hitchcock 2007, 503).

The idea is that an appropriate model is a sufficiently rich representation of causal reality that moving to a richer representation would not reveal an apparent actual causal relation to be spurious.[10]

The converse requirement – that a negative verdict about actual causation should not be overturned in a richer model – isn't needed. This is because actual causation (simpliciter) is defined in terms of actual causation relative to *at least one* appropriate model. A model relative verdict that $X = x$ is not an actual cause of $Y = y$ thus automatically fails to translate

[9] See Woodward (2005, 86–91).

[10] (Stability) renders the notion of an appropriate model relative to the causal claim being evaluated.

into a verdict that $X = x$ is not an actual cause (simpliciter) of $Y = y$ if there is a richer (and otherwise appropriate) model relative to which $X = x$ is an actual cause of $Y = y$.

We can now state a definition of actual causation in terms of appropriate PCMs that handles **(ProbPre)**.

## 6. PAC

Actual causation *simpliciter* is defined in terms of actual causation relative to an appropriate PCM. Model-relative actual causation is then defined.[11]

| **AC(S)** |
|---|
| *Where* $x, x' \in \mathscr{R}(X)$ *and* $y \in \mathscr{R}(Y)$, $X = x$ *(rather than* $X = x'$*) is an* actual cause (simpliciter) *of* $Y = y$ *in world* $\theta$ *iff* $X = x$ *(rather than* $X = x'$*) is an actual cause of* $Y = y$ *relative to* at least one *model* $\mathscr{M}$ *(with* $X, Y \in \mathscr{V}$*) that is appropriate for evaluating whether* $X = x$ *(rather than* $X = x'$*) is an* actual cause (simpliciter) *of* $Y = y$ *in* $\theta$. |

[11] Those familiar with Halpern and Pearl's (2001, 2005) analyses of DAC are invited to see an analogy with **AC(M-R)**. **AC(M-R)** was partly inspired by thinking about how a counterpart of Halpern and Pearl's analysis might be developed that is adequate to the probabilistic case. Ultimately, I'm optimistic that an adequate account of DAC will fall out of an adequate account of PAC as the special case where all probabilities are 1 or 0. This is why my definitions take the definiendum to be 'actual cause' rather than 'probabilistic actual cause'.

| **AC(M-R)** |
| --- |
| Where $x, x' \in \mathscr{R}(X)$ and $y \in \mathscr{R}(Y)$, $X = x$ (rather than $X = x'$) is an *actual cause* of Y = y relative to a model $\mathscr{M}$ (with $X, Y \in \mathscr{V}$) in world $\theta$ iff there is a partition $(\mathbf{Z}, \mathbf{W})$ of $\mathscr{V} \backslash X, Y$ and some setting $\mathbf{W} = \mathbf{w}'$ of the variables in $\mathbf{W}$ such that the $do(\cdot)$ function associated with $\mathscr{M}$ entails that, for all subsets $\mathbf{Z}'$ of $\mathbf{Z}$ (where, for each such subset, $\mathbf{Z}' = \mathbf{z}^*$ are the values that the variables in $\mathbf{Z}'$ have in $\theta$): |
| (**IN**) $P(Y = y \| do(X = x \& \mathbf{W} = \mathbf{w}' \& \mathbf{Z}' = \mathbf{z}^*)) > P(Y = y \| do(X = x' \& \mathbf{W} = \mathbf{w}'))$ |

**AC(M-R)** counts $M = 1$ as an actual cause of $E = 1$ relative to $\mathscr{M}_{Pre}$ (and the world described in (**ProbPre**)). Consider the partition of $\mathscr{V}_{Pre} \backslash M, E$ such that $\mathbf{W} = \{D, Y\}$ and $\mathbf{Z} = \{T\}$. And consider the assignment $\{D = 1, Y = 0\}$ of values to the variables in $\mathbf{W}$. **AC(M-R)** is satisfied because (**IN**) holds for both subsets of $\mathbf{Z}$ ($\emptyset$ and $\{T\}$), as shown by (4) and (5):

(4) $$P(E = 1|do(M = 1 \& D = 1 \& Y = 0)) > P(E = 1|do(M = 0 \& D = 1 \& Y = 0))$$

(5) $$P(E = 1|do(M = 1 \& T = 1 \& D = 1 \& Y = 0)) > P(E = 1|do(M = 0 \& D = 1 \& Y = 0))$$

Inequality (4) indicates that my action raises the probability of the explosion *under the contingency* – i.e. *holding fixed* – that (you make your decision but) don't place your Th-228 near the Geiger. The existence of this *contingent* probability-raising reflects the fact that there is a path – $\langle M, T, E \rangle$ – along which $M = 1$ promotes $E = 1$ (because $M = 1$ raises the probability of $E = 1$ when we hold fixed the values of all variables off that path). It is the existence of such a path – representing the process via which $M = 1$ produces $E = 1$ – that appears to drive our intuitions about actual causation in this case (cf. Hitchcock 2001).

Inequality (5) indicates that, again holding fixed $D = 1$ and $Y = 0$, the probability of $E = 1$ is higher if I place my U-232 near the Geiger *and the threshold reading is reached* than if I'd simply never placed my U-232 near the Geiger in the first place. As will be seen, this requirement ensures that, not only is there a potential process via which $M = 1$ threatens to bring about $E = 1$, but that process is complete.

Since **AC(M-R)** implies that $M = 1$ is an actual cause of $E = 1$ relative to $\mathscr{M}_{Pre}$, **AC(S)** yields the (correct) result that $M = 1$ is an actual cause (simpliciter) of $E = 1$ provided that $\mathscr{M}_{Pre}$ is appropriate. $\mathscr{M}_{Pre}$ *is* appropriate. Clearly it satisfies (Partition) and (Independence). It satisfies (Naturalness) because all of the states that its variables represent are reasonably natural. It was stipulated that the $do(\cdot)$ function associated with $\mathscr{M}_{Pre}$ is such that (Veridicality) is satisfied. $\mathscr{M}_{Pre}$ does not represent the sort of 'non-serious' possibility that (Serious Possibilities) is introduced to rule out (cf. Hitchcock 2001; Woodward 2005, 86–91).

Finally, (Stability) is satisfied because the causal process from my action to the explosion is complete. Holding fixed $Y = 0$, the probability of the explosion if $M = 1$ *and* part(s) of this process occur(s) is higher than the probability of the explosion if simply $M = 0$. Any variable (whose values represent reasonably natural states, form a partition, and are logically and metaphysically independent from the variables in $\mathscr{V}_{Pre}$) that might be added to $\mathscr{V}_{Pre}$ either represents part of this process or it doesn't. If it does, its actual value represents *the occurrence* of part of the process. So, if it is added to $\mathscr{V}_{Pre}$, including it in $\mathbf{Z}$ will not prevent (**IN**) from holding for all subsets $\mathbf{Z}'$ of $\mathbf{Z}$. If it doesn't, then adding it to $\mathscr{V}_{Pre}$, including it in $\mathbf{W}$, and holding it fixed at its actual value as part of the assignment $\mathbf{W} = \mathbf{w}'$ will not make a difference to the fact that (**IN**) holds for all subsets $\mathbf{Z}'$ of $\mathbf{Z}$, since holding fixed $Y = 0$ as part of $\mathbf{W} = \mathbf{w}'$ is already sufficient to ensure this.

**AC(M-R)** gives the verdict that $D = 1$ is *not* an actual cause of $E = 1$ relative to $\mathscr{M}_{Pre}$. Consider the partition of $\mathscr{V}_{Pre} \backslash D, E$ such that $\mathbf{W} = \{M\}$ and $\mathbf{Z} = \{Y, T\}$. Observe that:

(6) $$P(E = 1|do(D = 1\&M = 0)) > P(E = 1|do(D = 0\&M = 0))$$

And:

(7) $$P(E = 1|do(D = 1\&M = 1)) > P(E = 1|do(D = 0\&M = 1))$$

Thus, whichever possible value we hold fixed $M$ at, the probability of $E = 1$ is higher if $D = 1$ than if $D = 0$. So $D = 1$ contingently raises the probability of $E = 1$.[12] That's because there's a path – $\langle D, Y, E \rangle$ – along which $D = 1$ promotes $E = 1$.

**AC(M-R)** nevertheless entails that $D = 1$ is *not* an actual cause of $E = 1$ relative to $\mathscr{M}_{Pre}$. Consider the subset $\{Y\}$ of $\mathbf{Z}$, and observe that:

(8) $$P(E = 1|do(D = 1\&Y = 0\&M = 0)) \leq P(E = 1|do(D = 0\&M = 0))$$

And:

(9) $$P(E = 1|do(D = 1\&Y = 0\&M = 1)) \leq P(E = 1|do(D = 0\&M = 1))$$

That is, whichever possible value we hold fixed $M$ at, the probability of the explosion is no higher if you make your decision *but don't place your Th-228 near the Geiger* than if you'd

[12] The obtaining of just one of (6) or (7) would suffice to show this.

never made that decision in the first place. Thus (**IN**) does not hold for every subset of $\mathbf{Z}$ for this partition of variables no matter what values we assign to the variables in $\mathbf{W}$. This reflects the fact that, because you didn't place your Th-228 near the Geiger, there is no complete causal process by which your decision produces the explosion. Your non-placement of your Th-228 'neutralizes' the danger of your decision causing the explosion.

Is there an alternative partition ($\mathbf{W}$, $\mathbf{Z}$) of $\mathscr{V}_{Pre}$ and assignment $\mathbf{W} = \mathbf{w}'$ such that (**IN**) holds for all subsets $\mathbf{Z}'$ of $\mathbf{Z}$? (There need only be *one* for **AC(M-R)** to be satisfied.) There isn't. Assigning $Y$ to $\mathbf{W}$ instead of $\mathbf{Z}$ won't help, since the value of $Y$ 'screens off' $D$ from $E$. So, where $Y \in \mathbf{W}$, no assignment $\mathbf{W} = \mathbf{w}'$ will be such that, holding fixed $\mathbf{W} = \mathbf{w}'$, the probability of $E = 1$ is higher when $D = 1$ (and the variables in $\emptyset \subseteq \mathbf{Z}$ are set to their actual values) than when $D = 0$. So (**IN**) doesn't hold for all subsets $\mathbf{Z}'$ of $\mathbf{Z}$ for any such partition.

On the other hand, if we leave $Y$ in $\mathbf{Z}$ and also assign $M$ to $\mathbf{Z}$, then there are no variables in $\mathbf{W}$ to hold fixed. Now consider the subset $\{Y\}$ of $\mathbf{Z}$, and observe that:[13]

(10) $$P(E = 1|do(D = 1 \& Y = 0)) \leq P(E = 1|do(D = 0))$$

So, with $M$ assigned to $\mathbf{Z}$ it remains the case that (**IN**) doesn't hold for all subsets of $\mathbf{Z}$.

So there's no partition of $\mathscr{V}_{Pre} \backslash D, E$ such that (**IN**) is satisfied for all subsets of $\mathbf{Z}$ when we consider $D = 1$ as a putative cause of $E = 1$. **AC(M-R)** therefore doesn't count $D = 1$ as an actual cause of $E = 1$ relative to $\mathscr{M}_{Pre}$.

But for **AC(S)** to count $D = 1$ as an actual cause of $E = 1$ *simpliciter*, there need only be one appropriate model relative to which **AC(M-R)** counts $D = 1$ as an actual cause of $E = 1$. Is there such a model? There isn't. Suppose a candidate such model includes $Y$. Because $D$ is only relevant to $E$ because of its relevance to $Y$, the value of $Y$ 'screens off' the value of $D$

[13] Note: the fact that $Y = 0$ *due to an intervention* doesn't make $M = 1$ more likely.

from that of $E$. This means that, if $Y$ is included in $\mathbf{W}$ in the partition ($\mathbf{W}$, $\mathbf{Z}$) of the model's variable set and held fixed (either at 1 or 0) as part of the assignment $\mathbf{W} = \mathbf{w}'$, then (**IN**) won't be satisfied for the empty subset of $\mathbf{Z}$. Alternatively, if $Y$ is included in $\mathbf{Z}$ then, no matter what other variables are included in the model and assigned to $\mathbf{W}$, (**IN**) won't be satisfied for the subset $\{Y\}$ of $\mathbf{Z}$. Specifically, because $D = 1$ only threatens to bring about $E = 1$ because it threatens to bring about $Y = 1$, no matter what we hold fixed by inclusion on both sides of (**IN**), the probability of $E = 1$ is no higher if $D = 1$ *and* $Y = 0$ than if simply $D = 0$.

So **AC(M-R)** doesn't count $D = 1$ as an actual cause of $E = 1$ relative to any appropriate model with $Y$ in its variable set. This means that any otherwise appropriate model relative to which $D = 1$ *is* an actual cause of $E = 1$ can be expanded to a model in which $D = 1$ *isn't* an actual cause of $E = 1$ simply by the addition of $Y$. Provided the expanded model is appropriate, the original model violates (Stability) and is inappropriate. So **AC(S)** will correctly not count $D = 1$ as an actual cause *simpliciter* of $E = 1$.

Since the values of $Y$ form a partition and represent natural states of affairs, (Partition) and (Naturalness) will be satisfied by the expanded model if they were satisfied by the original model. With regard to (Veridicality), it should be noted that there are multiple ways of expanding the original model via the addition of $Y$, each associated with a different $do(\cdot)$ function from elements of $\mathscr{C}^*$ to probability distributions over $\mathscr{F}^*$ (where $\mathscr{C}^*$ and $\mathscr{F}^*$ are generated by the expanded variable set in the way described in Section 4). In looking for an apt expanded model, we just select the one with the $do(\cdot)$ function that returns the objective chances on $\mathscr{F}^*$ that *would* obtain as a result of interventions bringing about the various elements of $\mathscr{C}^*$. With regard to (Serious Possibilities) note that, given your decision, your placing *and* your not placing your Th-228 near the Geiger are both salient possibilities in

**(ProbPre)**. So it doesn't seem that the expanded model could represent any non-serious possibilities if the original model doesn't. (Independence) is a little trickier. Might not the original model include a variable whose values are logically or metaphysically related to those of $Y$? Given that the variables in the original model are assumed to satisfy (Partition) it seems that any variable logically or metaphysically related to $Y$ – e.g. $Y'$, which takes value $Y' = 0$ if you don't place your Th-228 near the Geiger, $Y' = 1$ if you place it 2.5-5cm from the Geiger, and $Y' = 2$ if you place it 0-2.5cm from the Geiger – will also be such that its actual value neutralizes the threat of $D = 1$ bringing about $E = 1$, so that **AC(M-R)** is not satisfied in the original model. The exception to this would be if the original model included a variable that represents a gerrymandered states of affairs – e.g. $Y''$, which takes value $Y'' = 1$ if you place your Th-228 near the Geiger *or* Obama is US president, and $Y'' = 0$ otherwise – in which case the original model will violate (Naturalness).

## 7. Conclusion

Drawing upon PCMs, an account of PAC has been given that gives a correct treatment of probabilistic preemption on intuitive grounds. Traditional accounts of PAC misdiagnose this central test case (Menzies, 1989, 1996; Hitchcock 2004). Examination of whether PCMs can help tackle some of the other outstanding problems of PAC is warranted.

## References

Anscombe, E. (1971). *Causality and Determination*. Cambridge: CUP.

Blanchard, T. and J. Schaffer (2016). Cause without Default. In H. Beebee, C. Hitchcock, and H. Price (Eds.), *Making a Difference*. Oxford: OUP.

Goldszmidt, M. and J. Pearl (1992). Rank-Based Systems. In *Proceedings of the Third International Conference on Knowledge Representation and Reasoning*, San Mateo, CA, pp. 661–672. Morgan Kaufmann.

Halpern, J. Y. (2014). Appropriate Causal Models and Stability of Causation. In *Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning*, Palo Alto, CA, pp. 198–207. AAAI Press.

Halpern, J. Y. and C. Hitchcock (2010). Actual Causation and the Art of Modeling. In R. Dechter, H. Geffner, and J. Y. Halpern (Eds.), *Heuristics, Probability and Causality*, pp. 383–406. London: College Publications.

Halpern, J. Y. and J. Pearl (2001). Causes and Explanations: A Structural-Model Approach. Part I: Causes. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, pp. 194–202. Morgan Kaufmann.

Halpern, J. Y. and J. Pearl (2005). Causes and Explanations: A Structural-Model Approach. Part I: Causes. *British Journal for the Philosophy of Science 56*, 843–87.

Hitchcock, C. (2001). The Intransitivity of Causation Revealed in Equations and Graphs. *Journal of Philosophy 98*, 194–202.

Hitchcock, C. (2004). Do All and Only Causes Raise the Probabilities of Effects? In J. Collins, N. Hall, and L. Paul (Eds.), *Causation and Counterfactuals*, pp. 403–417. Cambridge, MA: MIT Press.

Hitchcock, C. (2007). Prevention, Preemption, and the Principle of Sufficient Reason. *Philosophical Review 116*, 495–532.

Ismael, J. (2009). Probability in Deterministic Physics. *Journal of Philosophy 106*, 89–108.

Lewis, D. (1986). *Philosophical Papers*, Volume 2. Oxford: Oxford University Press.

Loewer, B. (2001). Determinism and Chance. *Studies in History and Philosophy of Science Part B 32*, 609–620.

Menzies, P. (1989). Probabilistic Causation and Causal Processes: A Critique of Lewis. *Philosophy of Science 56*, 642–663.

Menzies, P. (1996). Probabilistic Causation and the Pre-emption Problem. *Mind 105*, 85–117.

Paul, L. (2000). Aspect Causation. *Journal of Philosophy 97*, 235–256.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (Second ed.). Cambridge: CUP.

Reichenbach, H. (1971). *The Direction of Time*. Mineola, NY: Dover.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Suppes, P. (1970). *A Probabilistic Theory of Causality, Acta Philosophica Fennica*. Amsterdam: North-Holland.

Weslake, B. (2016). A Partial Theory of Actual Causation. Forthcoming in *British Journal for the Philosophy of Science*.

Woodward, J. (2005). *Making Things Happen: A Theory of Causal Explanation*. Oxford: OUP.