Causation as Production and Dependence

or, A Model-Invariant Theory of Causation

J. Dmitri Gallow

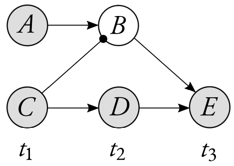
University of Pittsburgh, Center for the Philosophy of Science · February 13, 2018

1 Causal Models

1. We will represent causal determination structure with a *causal model*, or a *structural equations model*,

   Causal Models
   A causal model $\mathbb{M} = (\mathbb{U}, \vec{u}, \mathbb{V}, \mathbb{E}, \mathscr{D})$ is a 5-tuple of

   (a) A vector, $\mathbb{U} = (U_1, U_2, \dots, U_M)$, of *exogenous* variables;
   (b) An assignment of values, $\vec{u} = (u_1, u_2, \dots, u_M)$, to $\mathbb{U}$;
   (c) A vector $\mathbb{V} = (V_1, V_2, \dots, V_N)$, of *endogenous* variables;
   (d) A vector $\mathbb{E} = (\phi_{V_1}, \phi_{V_2}, \dots, \phi_{V_N})$ of *structural equations*, one for each $V_i \in \mathbb{V}$; and
   (e) A specification, $\mathscr{D}$, of which variable values are *default* and which are *deviant*.



$$\mathbb{U} : (A, C) \qquad \vec{u} : (1, 1) \qquad \mathbb{V} : (B, D, E)$$

$$\mathbb{E} : \begin{pmatrix} E := B \vee D \\ D := C \\ B := A \wedge \neg C \end{pmatrix}$$

Figure 1: *Preemptive Overdetermination*. (For all variables, the value 0 is default, and the value 1 is deviant.)

   (a) Given a neuron diagram, let the *canonical model* be the one that has, for each neuron, a binary variable taking the value 1 if the neuron fires and the value 0 if it doesn't fire (where not firing is default, firing deviant), and a true system of equations describing how the values of those variables are causally determined by each other. I'll assume throughout that the canonical model of a neuron diagram is correct.

2. Given a causal model $\mathbb{M}$, and an assignment $\mathbf{v}$ of values to the variables in $\mathbf{V}$, we can define a *counterfactual model* $\mathbb{M}[\mathbf{V} \to \mathbf{v}]$.

   Counterfactual Causal Models
   Given a causal model $\mathbb{M} = (\mathbb{U}, \vec{u}, \mathbb{V}, \mathbb{E})$, including the variables $\mathbf{V}$, and given the assignment of values $\mathbf{v}$ to $\mathbf{V}$, the counterfactual model $\mathbb{M}[\mathbf{V} \to \mathbf{v}] = (\mathbb{U}[\mathbf{V} \to \mathbf{v}], \vec{u}[\mathbf{V} \to \mathbf{v}], \mathbb{V}[\mathbf{V} \to \mathbf{v}], \mathbb{E}[\mathbf{V} \to \mathbf{v}], \mathscr{D}[\mathbf{V} \to \mathbf{v}])$ is the model such that:
   (a) $\mathbb{U}[\mathbf{V} \to \mathbf{v}] = \mathbb{U} \cup \mathbf{V}$
   (b) $\vec{u}[\mathbf{V} \to \mathbf{v}] = \vec{u} \cup \mathbf{v}$
   (c) $\mathbb{V}[\mathbf{V} \to \mathbf{v}] = \mathbb{V} - \mathbf{V}$
   (d) $\mathbb{E}[\mathbf{V} \to \mathbf{v}] = \mathbb{E} - (\phi_{V_i} \mid V_i \in \mathbf{V})$
   (e) $\mathscr{D}[\mathbf{V} \to \mathbf{v}] = \mathscr{D}$

3. Using counterfactual models, we may provide a semantics for causal counterfactuals:

> Causal Counterfactuals
> In a causal model $\mathbb{M}$, containing the variables in $\mathbf{V}$, the causal counterfactual $\mathbf{V} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} \psi$ is true iff ψ is true in the counterfactual model $\mathbb{M}[\mathbf{V} \to \mathbf{v}]$,
>
> $$\mathbb{M} \models \mathbf{V} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} \psi \iff \mathbb{M}[\mathbf{V} \to \mathbf{v}] \models \psi$$

2 Model Invariance

4. Ideally, a theory of causation would satisfy the following principle:

   Model Invariance Given any two causal models, $\mathbb{M}$ and $\mathbb{M}^\dagger$, which both contain the variables C and E, if both $\mathbb{M}$ and $\mathbb{M}^\dagger$ are correct, then $C = c$ caused $E = e$ in $\mathbb{M}$ iff $C = c$ caused $E = e$ in $\mathbb{M}^\dagger$.

5. In general, if $\mathbb{M} = (\mathbb{U}, \vec{u}, \mathbb{V}, \mathbb{E}, \mathscr{D})$ is a causal model with $U \in \mathbb{U}$, then let $\mathbb{M}^{-U}$ be the model that you get by:

   (a) Removing U from $\mathbb{U}$
   (b) Removing U's value from $\vec{u}$
   (c) Exogenizing any variables in $\mathbb{V}$ whose only parent was U
   (d) Replacing U for its value in every structural equation in $\mathbb{E}$
   (e) Removing default information about U from $\mathscr{D}$.

6. If every equation in $\mathbb{M}^{-U}$ is surjective, then say that U is an *inessential* variable. Then, we should endorse the following principle:

   Exogenous Reduction If a causal model $\mathbb{M} = (\mathbb{U}, \vec{u}, \mathbb{V}, \mathbb{E}, \mathscr{D})$ is correct, and $U \in \mathbb{U}$ is inessential, then $\mathbb{M}^{-U}$ is also correct.

7. In general, if $\mathbb{M} = (\mathbb{U}, \vec{u}, \mathbb{V}, \mathbb{E}, \mathscr{D})$ is a causal model with $V \in \mathbb{V}$, then let $\mathbb{M}^{-V}$ be the model that you get by:

   (a) Leaving $\mathbb{U}$ alone
   (b) Leaving $\vec{u}$ alone
   (c) Removing V from $\mathbb{V}$
   (d) Removing ϕ_V from $\mathbb{E}$, and replacing V with $\phi_V(\mathbf{PA}(V))$ wherever V appears on the right-hand-side of an equation in $\mathbb{E}$[1]
   (e) Removing default information about V from $\mathscr{D}$

8. (a) If V has a single parent, Pa, and a single child, Ch, and if Pa is not *also* a parent of Ch, then say that V is an *interpolated* variable.

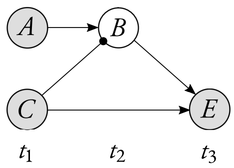
   $$\ldots Pa \to V \to Ch \ldots$$

   (b) If V is interpolated and all the equations in $\mathbb{M}^{-V}$ are surjective, then say that V is *inessential*.
   (c) Then, we should accept the following principle:

   Endogenous Reduction If a causal model $\mathbb{M} = (\mathbb{U}, \vec{u}, \mathbb{V}, \mathbb{E}, \mathscr{D})$ is correct, and $V \in \mathbb{V}$ is an inessential variable, then $\mathbb{M}^{-V}$ is also correct.

9. Though there isn't the space to show it here, the accounts of Hitchcock (2001, 2007), Halpern & Pearl (2001, 2005), Woodward (2003), Halpern (2008), and Weslake (forthcoming) are all inconsistent with **Model Invariance**, **Exogenous Reduction**, and **Endogenous Reduction**.

[1] $\mathbf{PA}(E)$ are E's *causal parents* in the model—those variables which appear on the right-hand-side of E's structural equation ϕ_E.



$$\mathbb{U} : (A, C) \qquad \vec{u} : (1, 1) \qquad \mathbb{V} : (B, E) \qquad \mathbb{E} : \begin{pmatrix} E \coloneqq B \vee C \\ B \coloneqq A \wedge \neg C \end{pmatrix}$$

Figure 2: *Preemptive Overdetermination*

3 A Model Invariant Theory of Causation

10. I will present a theory of causation, formulated within the framework of structural equations models, which is consistent with **Endogenous Reduction**, **Exogenous Reduction**, and **Model Invariance**.

    (a) I'll build up the theory by progressing through some familiar cases from the literature.

3.1 Preemptive Overdetermination

11. (a) In the canonical model, $\mathbb{M}_2$, of *Preemptive Overdetermination* shown in figure 2, $E = 1$ does not counterfactually depend upon $C = 1$.

    (b) *However*, if we just look at E's structural equation $E \coloneqq B \vee C$, and B and C's actual values, then $E = 1$ *does* counterfactually depend upon $C = 1$. Call this submodel of $\mathbb{M}_2$ the *local model* at E.

12. In general, we can define the *local model at E* as follows.

    Local Causal Model
    Given a causal model $\mathbb{M} = (\mathbb{U}, \vec{u}, \mathbb{V}, \mathbb{E}, \mathscr{D})$, with $E \in \mathbb{V}$, the *local model* at E, $\mathbb{M}(\!(E)\!)$, is the causal model in which

    (a) The exogenous variables are just the parents of E, $\mathbf{PA}(E)$, in the original model $\mathbb{M}$;
    (b) The exogenous variables $\mathbf{PA}(E)$ are assigned the values they take on in $\mathbb{M}$;
    (c) The sole endogenous variable is E;
    (d) The sole structural equation is E's structural equation in $\mathbb{M}$, ϕ_E; and
    (e) The defaults for E and $\mathbf{PA}(E)$ are the same as in $\mathbb{M}$.

13. Say that $E = e$ *locally* counterfactually depends upon $C = c$ iff, *in the local model at* E, $\mathbb{M}(\!(E)\!)$, there's some c^*, e^* such that

$$\mathbb{M}(\!(E)\!) \models C = c^* \;\Box\!\!\rightarrow E = e^*$$

14. A (preliminary) proposal, then, is that *either* local *or* global counterfactual dependence suffices for causation.

    (a) While this helps with the case of preemptive overdetermination in figure 2, it does nothing to help with the neuron diagram from figure 1.
    (b) It would be nice to handle that case by appealing to the transitivity of causation.
    (c) Unfortunately, there are a number of counterexamples to the transitivity of causation.

3.2 Counterexamples to Transitivity

15. Sometimes, we can trace out of sequence of causal relations and conclude that the event at the start of the chain caused the one at the end. If that's so, then I'll call the chain *transitive*.

Figure 3: *Tampering* (*cf.* PAUL & HALL 2013). The octogonal neurons can either fire weakly (light grey) or strongly (dark grey). If C fires, this diminishes the strength with which B fires. In figure 3(a), C's firing caused B to fire weakly. And B's firing weakly caused E to fire. But C's firing didn't cause E to fire.

(a) LEWIS thought that causal chains were always transitive, but this has unpalatable consequences. Chris smokes, contracts cancer, undergoes chemo, and survives. The smoking causes the cancer; the cancer causes the chemo; and the chemo causes the survival—so LEWIS is forced to say that the smoking causes the survival.

(b) The right thing to say is that causal chains are sometimes, but not always, transitive. The difficulty is working out just *when*.

16. The plan: I'll attempt to give conditions specifying when a directed path, **P**, running from the variable V_1 to the variable V_N,

$$\mathbf{P} = V_1 \to V_2 \to V_3 \to \cdots \to V_N$$

permits the inference that $V_1 = v_1$ caused $V_N = v_N$. When it does, I'll call the path a *transitive path*.

17. One kind of counterexample to transitivity is illustrated by the neuron diagram in figure 3. C's firing caused B to fire weakly (rather than strongly); B's firing weakly (rather than not) caused E to fire. But C's firing didn't cause E to fire.[2]

(a) The solution: adopt a *contrastivist* theory of causation, and require that the *contrasts* in our causal chain match up.[3]

(b) Note: once we go contrastivist, we will be theorizing in terms of a 4-place causal relation

$$\text{CAUSE}(C = c, C = c^*, E = e, E = e^*)$$

From this, we may recover a familiar 2-place causal relation:

$$\text{CAUSE}(C = c, E = e) \iff \exists c^* \exists e^* \text{CAUSE}(C = c, C = c^*, E = e, E = e^*)$$

18. For two other counterexamples to transitivity, consider the neuron diagrams in figure 4.

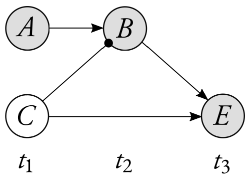
(a) In both cases, either the start or the end of the causal chain involves a *default* variable value.

(b) This suggests the hypothesis: in order for a directed path to be a *transitive* path, the variable values at the start and end of that path must both be *deviant* (and, though I won't be motivating this requirement here, their contrasts must also be *default*).

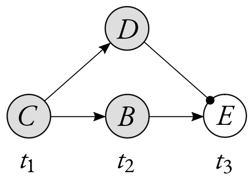
19. In general, this will be our account of which a directed path in a causal model is *transitive*:

[2] *Cf.* MCDERMOTT (1995)'s *Dog Bite* example and the counterexamples to transitivity discussed in PAUL (2004).

[3] *cf.* SCHAFFER (2005).



(a)



(b)

Figure 4: In figure 4(a), C's failure to fire causes B to fire. B's firing causes E to fire. But C's failure to fire doesn't cause E to fire. In figure 4(b), C's firing causes D to fire. D's firing causes E to remain dormant. But C's firing does not cause E to remain dormant.

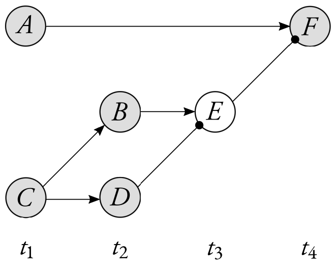


Figure 5: Short Circuit, again. C's firing didn't cause F to fire, but without condition (c) of Transitive Path, we would have to say that it did.

Transitive Path
In a causal model $\mathbb{M}$, a directed path running from V_1 to V_N

$$\mathbf{P} = V_1 \to V_2 \to V_3 \to \cdots \to V_N$$

is a *transitive path* iff:

(a) For each variable V_i along $\mathbf{P}$, there is a pair (v_i, v_i^*) of V_i's actual value v_i in $\mathbb{M}$, and a *contrast* value v_i^*,

$$(v_1, v_1^*) \to (v_2, v_2^*) \to (v_3, v_3^*) \to \cdots \to (v_N, v_N^*)$$

such that: for all j between 1 and $N-1$, V_j's taking on the value v_j, rather than v_j^*, caused V_{j+1} to take on the value v_{j+1}, rather than v_{j+1}^*;

(b) Both V_1's and V_N's actual values are *deviant*, their contrast values *default*; and

(c) Every *departure* variable along $\mathbf{P}$ causes each of its *return* variables along $\mathbf{P}$.[4]

(a) To see the reason for this final condition, consider the neuron diagram in figure 5.

3.3 Prevention and Omission without Dependence?

20. So far, we've only looked at causal relations where both the cause and effect variables take on deviant values. But default variable values can also be causes and effects.

(a) Because counterfactual dependence suffices for causation, cases of *prevention* (figure 6) and *omission* (figure 7) involve default effects and causes, respectively.

[4] For any variables D, R along the path (unless $(D,R) = (V_1, V_N)$), D is a *departure* variable, and R is one of its *return* variables iff there is a path, $\mathbf{O} = D \to O_1 \to O_2 \to \cdots \to R$, such that D and R are the only variables from $\mathbf{P}$ on $\mathbf{O}$.

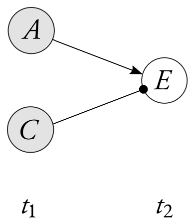


Figure 6: *Prevention*

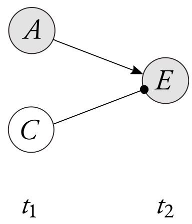


Figure 7: *Omission*

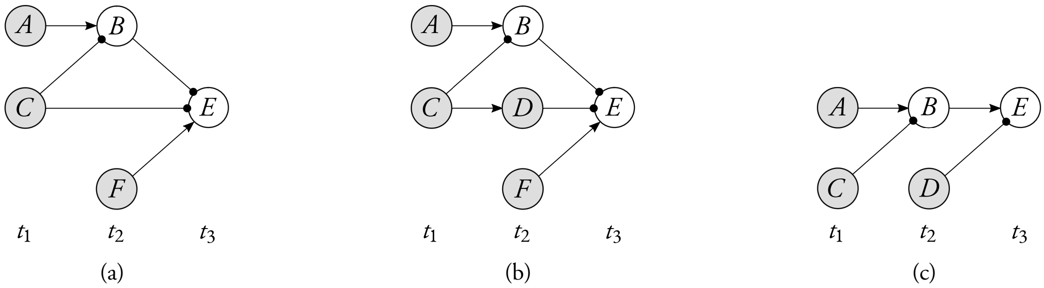


Figure 8: Prevention without Dependence?

21. When $C = c$ and $E = e$ were *deviant* variable values, we said that *local* counterfactual dependence was sufficient for causation. Should we say the same thing when $C = c$ or $E = e$ is a *default* variable value?

    (a) This question turns out to be closely related to cases of *Preemptive Prevention* (or, cases of prevention without dependence) like the one shown in figure 8.

    (b) If we say that local counterfactual dependence suffices for causation, then, in the canonical model of the neuron diagram in figure 8(a), we will say that C's firing caused E to not fire.

    (c) However, we would not be able to say the same thing about the neuron diagram in figure 8(b). For, in the canonical model of that neuron diagram, $E = 0$ does not *locally* counterfactually depend upon $C = 1$, since C isn't even in the local model at E. Moreover, since E's remaining dormant is a *default* state of that neuron, we would not be able to appeal to the transitivity condition to say that C's firing prevented E from firing.

    (d) So, we should say that, in the case where the cause or effect variable value is default, local counterfactual dependence is *not* sufficient for causation, and therefore, in the cases shown in figure 8, C's firing *doesn't* prevent E from firing all by itself.[5]

22. We can further support this treatment by noting that, if we want a model-invariant account of causation, then we are *forced* to say, in figure 8(c), that C's firing prevented E from firing iff D's firing *also* prevented E from firing.

    (a) Beginning with the canonical causal model of figure 8(c), **Exogenous Reduction** allows us to remove the inessential exogenous variable A from our model. Then, **Endogenous Reduction** allows us to remove the inessential interpolated variable B. We end up with a causal model containing the sole structural equation $E := \neg C \wedge \neg D$. But this equation treats C and D symmetrically, and both C and D take on the same value. So, any account of causation will say that, in this model, $C = 1$ caused $E = 0$ iff $D = 1$ caused $E = 0$. Since $D = 1$ clearly did *not* cause $E = 1$, any model-invariant account of causation should say that $C = 1$ didn't caused $E = 0$ either.

23. So there is no prevention without dependence. Can there be omission without dependence? For parallel reasons, it does not appear so. Consider the neuron diagrams from figure 9.

[5] We can still say that the *disjunction* of A's firing and C's firing caused E to remain dormant.

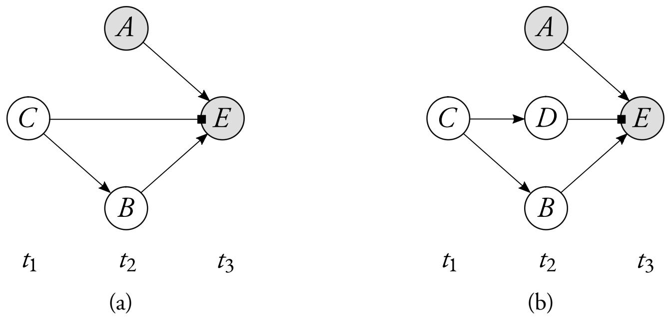


Figure 9: Omission without Dependence?

(a) Suppose we said that local dependence sufficed for $C = 0$ causing $E = 1$ in figure 9(a). Then, we should say the same thing about figure 9(b). However, in the canonical model of figure 9(b), there is no local dependence between $E = 1$ and $C = 0$. Moreover, there can be no transitive path running from C to E in that canonical model, since C's actual value is default.

24. Exactly similar issues arise when we consider cases of local dependence where both C and E's values are default (left as an exercise). So we should conclude that local dependence suffices for causation *only when* C and E both have deviant values (and default contrasts).

3.4 Causation as Production and Dependence

25. In summary, we have arrived at the following account of causation:

Causation as Production and Dependence
In a causal model $\mathbb{M}$, C's taking on the value c, rather than c^*, caused E to take on the value e, rather than e^*, iff either (Prod) or (Dep).

(Prod) Both c and e are *deviant* variable values, the contrasts c^* and e^* *defaults*, and either:

i. In the local model at E, $\mathbb{M}(\!(E)\!)$, had C taken on the value c^*, E would have taken on the value e^*,

$$M(\!(E)\!) \models C = c^* \;\Box\!\!\rightarrow E = e^*$$

or

ii. In $\mathbb{M}$, there is a transitive path leading from C to E.

(Dep) In $\mathbb{M}$, had C taken on the value c^*, E would have taken on the value e^*,

$$\mathbb{M} \models C = c^* \;\Box\!\!\rightarrow E = e^*$$

26. This account is consistent with **Model Invariance**, **Exogenous Reduction**, and **Endogenous Reduction**. Suppose that we have a correct model $\mathbb{M} = (\mathbb{U}, \vec{u}, \mathbb{V}, \mathbb{E}, \mathscr{D})$, with $U \in \mathbb{U}$ and $V \in \mathbb{V}$. And suppose that neither U nor V are C or E, and both U and V are inessential. Then:

(a) If $C = c$ caused $E = e$ in $\mathbb{M}$, then $C = c$ caused $E = e$ in $\mathbb{M}^{-U}$;

(b) If $C = c$ caused $E = e$ in $\mathbb{M}$, then $C = c$ caused $E = e$ in $\mathbb{M}^{-V}$;

(c) If $C = c$ *didn't* cause $E = e$ in $\mathbb{M}$, then $C = c$ *didn't* cause $E = e$ in $\mathbb{M}^{-U}$; and

(d) If $C = c$ *didn't* cause $E = e$ in $\mathbb{M}$, then $C = c$ *didn't* cause $E = e$ in $\mathbb{M}^{-V}$.

27. (a) The clause (Prod), taken in isolation, does a reasonably good job of capturing a notion of *causal production*. According to it, production involves the local, uninterrupted propagation of deviant, non-inertial states of affairs (rather than default, inertial states of affairs).

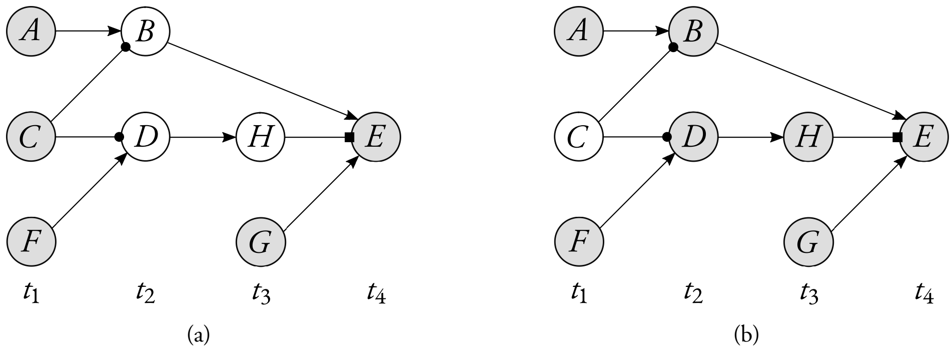


Figure 10: Double Prevention without Dependence. Figure 10(b) shows what would have happened, had C not fired, in figure 10(a).

(b) A hypothesis: the notion of causal production encapsulated in the Production clause of this account represents the core of our concept of causation. Those causal judgments which are licensed by the Production clause alone are far more intuitive, natural, and widespread than those which are only licensed with the addition of the Dependence clause.

   i. For instance, the Production clause is all that is required to show that C's firing caused E to fire in cases of *Preemptive Overdetermination*. And the judgment that C caused E in this case is widespread and uncontested.

   ii. In contrast, in order to establish causation in cases of prevention, omission, omissive prevention, double prevention, and so on, we will need to appeal to the Dependence clause. And these judgments are all less uniform and more controversial.

(c) However, if we accept the Production clause, then the full strength of the Dependence clause is required, if we are to satisfy **Model Invariance**, **Exogenous Reduction**, and **Endogenous Reduction**.

   i. Consider, for instance, the neuron diagram shown in figure 10(a). I take it that it is far from clear what to say about whether C's firing caused E to fire in figure 10(a). But, by removing inessential variables, this neuron diagram may be reduced to the model of preemptive overdetermination from figure 1.

   ii. In the case of preemptive overdetermination, we must say that $C = 1$ caused $E = 1$ (by Prod). So, if we wish our account to be model-invariant, then we must say that $C = 1$ caused $E = 1$ in the canonical model of figure 10(a).

   iii. But in order to conclude this with the transitivity clause, we must have $C = 1$ cause $D = 0$, $D = 0$ cause $H = 0$, and $H = 0$ cause $E = 1$.

   iv. So, we must count as causal cases of prevention, omission, and omissive prevention. We must appeal to the full strength of Dependence.

References

Collins, John, Ned Hall & L. A. Paul, editors. 2004. *Causation and Counterfactuals*. The MIT Press, Cambridge, MA. [9]

Halpern, Joseph Y. 2008. "Defaults and Normality in Causal Structures." *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning*, 198–208. [2]

—. 2016. *Actual Causality*. MIT Press, Cambridge, MA.

Halpern, Joseph Y. & Christopher Hitchcock. 2010. "Actual Causation and the Art of Modeling." In *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, Rina Dechter, Hechtor Geffner & Joseph Y. Halpern, editors, 383–406. College Publications.

—. 2015. "Graded Causation and Defaults." *The British Journal for the Philosophy of Science*, vol. 66 (2): 413–457.

Halpern, Joseph Y. & Judea Pearl. 2001. "Causes and Explanations: A Structural-Model Approach. Part 1: Causes." In *Proceedings of the Seventeeth Conference on Uncertainty in Artificial Intelligence*, John Breese & Daphne Koller, editors, 194–202. Morgan Kaufman, San Francisco. [2]

—. 2005. "Causes and Explanations: A Structural-Model Approach. Part 1: Causes." *The British Journal for the Philosophy of Science*, vol. 56: 843–887. [2]

Hitchcock, Christopher. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs." *The Journal of Philosophy*, vol. 98 (6): 273–299. [2]

—. 2007. "Prevention, Preemption, and the Principle of Sufficient Reason." *Philosophical Review*, vol. 116 (4): 495–532. [2]

Lewis, David K. 2000. "Causation as Influence." *The Journal of Philosophy*, vol. 97 (4): 182–197. Reprinted in Collins et al. (2004, pp. 75–106). [4]

—. 2004. "Causation as Influence." In Collins et al. (2004), chap. 3, 75–106. [4]

McDermott, Michael. 1995. "Redundant Causation." *The British Journal for the Philosophy of Science*, vol. 46 (4): 523–544. [4]

McGrath, Sarah. 2005. "Causation by Omission: A Dilemma." *Philosophical Studies*, vol. 123: 125–148.

Paul, L. A. 2004. "Aspect Causation." In Collins et al. (2004). [4]

Paul, L. A. & Ned Hall. 2013. *Causation: A User's Guide*. Oxford University Press, Oxford. [4]

Schaffer, Jonathan. 2005. "Contrastive Causation." *The Philosophical Review*, vol. 114 (3): 297–328. [4]

Weslake, Brad. forthcoming. "A Partial Theory of Actual Causation." *The British Journal for the Philosophy of Science*. [2]

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford. [2]